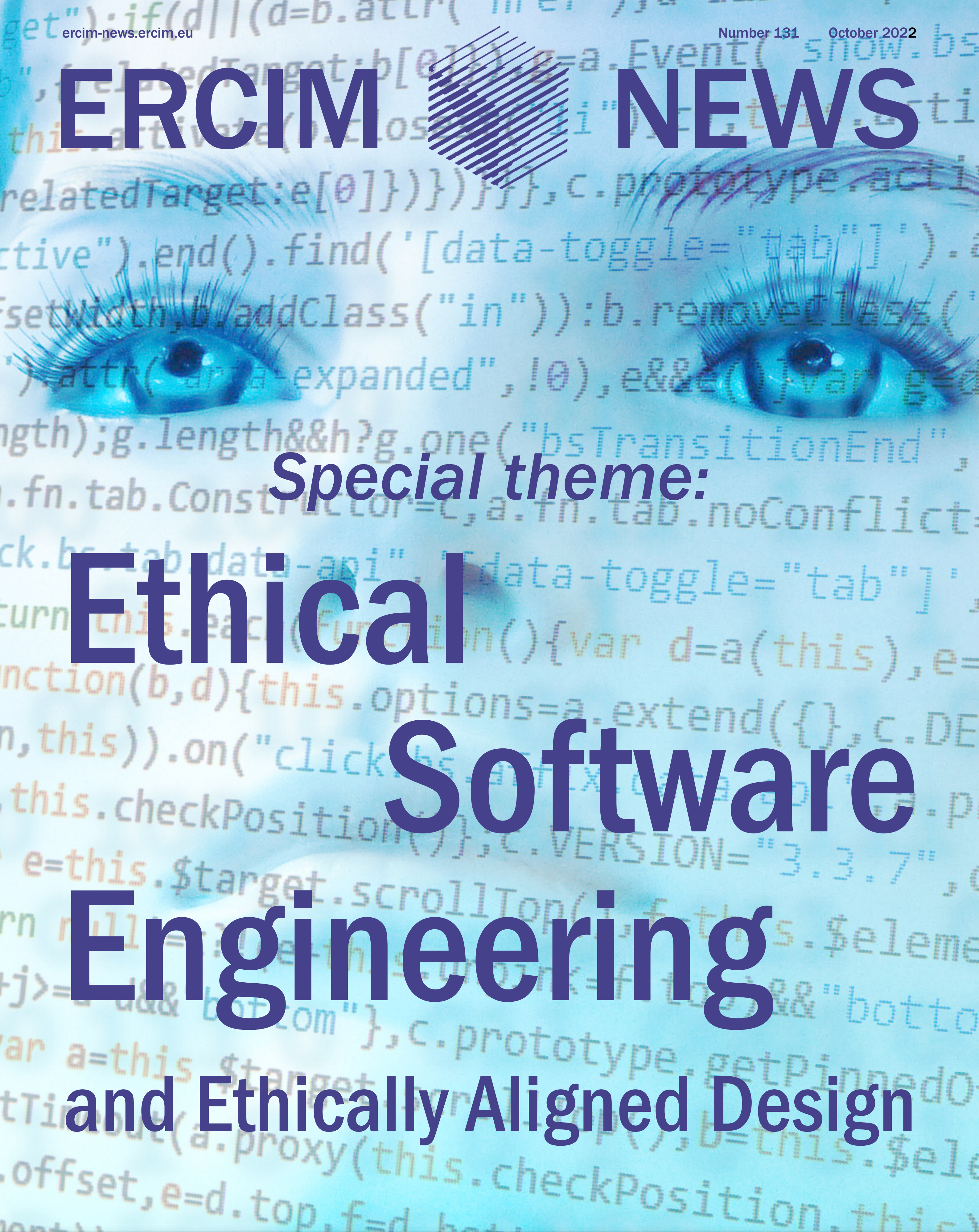
ercim-news.ercim.eu

Number 131 October 2022

ERCIM NEWS



Special theme:

Ethical Software Engineering and Ethically Aligned Design

Also in this issue

Research and Innovation:

Plausible Reasoning that Mimics Human Argumentation

Editorial Information

ERCIM News is the magazine of ERCIM. Published quarterly, it reports on joint actions of the ERCIM partners, and aims to reflect the contribution made by ERCIM to the European Community in Information Technology and Applied Mathematics. Through short articles and news items, it provides a forum for the exchange of information between the institutes and also with the wider scientific community. This issue has a circulation of about 2,000 printed copies and is also available online ,at https://ercim-news@ercim.eu.

ERCIM News is published by ERCIM EEIG
BP 93, F-06902 Sophia Antipolis Cedex, France
+33 4 9238 5010, contact@ercim.eu
Director: Philipp Hoschka, ISSN 0926-4981

Contributions
Contributions should be submitted to the local editor of your country

ERCIM News online edition: *https://ercim-news.ercim.eu/*

Next issue:
January 2023: Cognitive AI & Cobots

Subscription
Subscribe to ERCIM News by sending an email to en-subscriptions@ercim.eu

Cover illustration by Gerd Altmann from Pixabay.

JOINT ERCIM ACTIONS

SPECIAL THEME

The special theme "Ethical Software Engineering and Ethically Aligned Design" has been coordinated by the guest editors Georgia Kapitsaki (University of Cyprus) and Erwin Schoitsch (Austrian Institute of Technology AIT)

RESARCH AND INNOVATION

ANNOUNCEMENTS / IN BRIEF

Fabio Carrara is the Winner of the 2022 ERCIM Cor Baayen Young Researcher Award

The ERCIM Evaluation Committee for the Cor Baayen Young Researcher Award unanimously selected Fabio Carrara from ISTI-CNR as the winner for 2022. An honorary mention was given to Cédric Colas from Inria.

Fabio Carrara is currently employed at the Information Science and Technologies Institute, National Research Council (ISTI-CNR). He obtained his PhD in 2019 from the University of Pisa, Italy. The title of his thesis is "Deep Learning for Image Classification and Retrieval: Analysis and Solutions to Current Limitations". The work was supervised by Prof. Giuseppe Amato, Prof. Claudio Gennaro, and Prof. Francesco Marcelloni.

Fabio's research focuses mainly on deep learning for multimedia understanding, representation, and retrieval. Over the years, he has contributed to these areas of research from a theoretical and an applied perspective.

His professional experience encompasses collaborations with national and international institutions, such as the Masaryk University (Czech Republic), Scuola Normale Superiore (Italy), Institute for Systems and Robotics (Portugal), and several Italian CNR institutes (IN-CNR, ILC-CNR, IIT-CNR). Fabio has been involved in several European and Italian research projects (e.g., AI4Media, CNR4C-AIMAP, ADA, Smart News).

He has published more than 40 papers in international journals (e.g., Medical Image Analysis, IEEE TIP, Computer

Vision and Image Understanding, Information Systems, Information Processing & Management, MTAP, ESWA) and conferences in the areas of deep learning, computer vision, and multimedia information retrieval. He is an active member of the scientific community, and he has a good track record as a reviewer of international conferences and journals.

In 2018 he won the ISTI Young Research Award for best young (under 32 years) researcher.

Fabio's research activity stands out for its quality and interdisciplinarity. Worthy of note is his work on adversarial attack detection, proposing solutions based on the analysis of the features extracted by the various layers of deep neural networks. He also researched the application of appropriately simplified deep neural networks on resource-constrained devices, such as smart cameras. His research results are not only theoretical but also have significant application and technology transfer implications, as for example, the miniaturised models for parking occupancy detection (http://cnrpark.it/).

Cor Baayen Young Researcher Award 2022

Winner:
- Fabio Carrara (Information Science and Technologies Institute, National Research Council (ISTI-CNR)), nominated by Giuseppe Amato (CNR)

Honorary mention:
- Cédric Colas (Inria), nominated by Pierre-Yves Oudeyer (Inria)

Finalists:
- Simon Bibri, nominated by John Krogstie (NTNU);
- Gabrielle De Micheli (University of California, San Diego), nominated by Jean-Frédéric Gerbeau (Inria)
- Joao Gante (Hugging Face), nominated by Leonel Sousa (INESC)
- Toni Heittola (Tampere University), nominated by Katja Ojakangas (VTT)
- Johannes Mueller-Roemer (Fraunhofer IGD), nominated by Arjan Kuijper (Fraunhofer-Gesellschaft)
- Antonis Papaioannou (ICS-FORTH), nominated by Dimitris Plexousakis (ICS-FORTH)
- Shazia Tabassum (INESC TEC), nominated by João Gama (INESC TEC).

Evaluation Committee:
The Evaluation Committee constituents were Monica Divitini (NTNU – chair of the ERCIM Human Capital Task Group), Krzysztof Apt (CWI), Gabriel David (INESC), Georgia Kapitsaki (University of Cyprus), Thierry Priol (Inria), Fabrizio Sebastiani (ISTI-CNR), Jerzy Tiuryn (University of Warsaw). The decision was unanimous.

More information about the ERCIM Cor Baayen Young Researcher Award:
https://www.ercim.eu/human-capital/cor-baayen-award

27th International Conference on Formal Methods for Industrial Critical Systems - FMICS'22

by Maurice ter Beek

The yearly conference of the ERCIM Working Group on Formal Methods for Industrial Critical Systems, FMICS, the key conference at the intersection of industrial applications and formal methods, reached its 27th edition this year. Moreover, after two years online, this year's participants met in Warsaw, Poland, on 14 and 15 September.

The aim of the FMICS conference series is to provide a forum for researchers interested in the development and application of formal methods in industry. It strives to promote research and development for the improvement of formal methods and tools for industrial applications.

The conference was chaired by Marieke Huisman (University of Twente, The Netherlands) and Jan Friso Groote (Eindhoven University of Technology, The Netherlands) and organised under the umbrella of CONFEST 2022, alongside with CONCUR, QEST and FORMATS, organised by the general chair Sławek Lasota (University of Warsaw, Poland) and his team. FMICS 2022 attracted participants from many countries worldwide, both from academia and industry.

The international program committee, with members from 16 different countries, received 22 submissions of authors from 12 different countries, and decided to accept 13 papers after a thorough reviewing process. The program moreover included two excellent invited keynote presentations, namely "Reinforcement Learning with Guarantees that Hold for Ever" by Sven Schewe (Liverpool University, UK) and "Supporting Railway Innovations with Formal Modelling and Verification" by Bas Luttik (Eindhoven University of Technology, The Netherlands).

Following a tradition established over the years, Springer sponsored an award for the best FMICS paper. This year, the program committee selected the contribution "Deductive Verification of Smart Contracts with Dafny" by Franck Cassez, Joanne Fuller, Horacio Mijail, and Antón Quiles for the FMICS 2022 Best Paper Award.

FMICS 2023 will take place in Antwerp, Belgium, again under the umbrella of CONFEST 2023.

Links:
FMICS 2022 conference website:
https://fmics2022.fsa.win.tue.nl/

ERCIM WG FMICS: https://fmics.inria.fr/

Reference:
[1] J.F. Groote and M. Huisman (eds.). Formal Methods for Industrial Critical Systems: Proc. of the 27th Int. Conf. on Formal Methods for Industrial Critical Systems (FMICS'22). Springer LNCS, vol. 13487, 2022.
https://doi.org/10.1007/978-3-031-15008-1

Please contact:
Maurice ter Beek
ISTI-CNR, Italy
maurice.terbeek@isti.cnr.it

Franck Cassez receives the FMICS 2022 Best Paper Award from Marieke Huisman (picture by Paula Herber).

Introduction to the Special Theme

Ethical Software Engineering and Ethically Aligned Design

by Georgia Kapitsaki (University of Cyprus) and Erwin Schoitsch (AIT Austrian Institute of Technology)

Software creation makes our daily lives easier, tackling different aspects of our daily activities (e.g., work, leisure, mobility, communication, smart homes, digital governance of public authorities, etc.). Applications are linked with several ethical issues and require ethically aligned design, with software as key factor, covering a wide range of domains, from healthcare to military systems. In order to fully benefit from any type of application, the alignment of software applications, embedded into systems and systems-of-systems, with ethical values and principles is of tremendous importance. Constructing software and systems in an ethical way is linked with many specific societal values, including individuality, equality, and democracy. Approaches and mechanisms that assist in producing ethics-aware software and systems are vital, in order to protect the main rights of the end-users and all other affected entities.

Figure 1: IEEE Ethical aligned Design (2019) and UNESCO Recommendations (Nov. 2021).

According to the Stack Overflow Developer Survey in 2018, “Developers are not sure how they would report ethical problems and have differing ideas about who ultimately is responsible for unethical code.”

Ethical aspects are becoming a key challenge particularly in Europe, based on European values with respect to human rights, human independence, freedom of decision and self-determination, democracy, and privacy. This results in the requirement raised in recent EC research projects (Horizon Europe, KDT Key Digital Technologies) to not only take care of standardisation, but also ethical and societal aspects more than before, by, for example, including an “Ethical Advisor” as a role or task. This was also the case for example, in recent Austrian national-funded projects; for example, the guest editor Erwin Schoitsch took over this role in a few European (AI4CSM, FOCETA, AI-IQ Ready) and national projects (ADEX – Autonomous Driving Examiner, Austrian Research Promotion Agency FFG/BMK).

Several informatics standardisation organisations, national Ethics Commissions, and UNESCO have published ethics guidelines, demonstrating the rising awareness on a global level, most of them focusing on AI and decision-making, but also on general issues like the requirement of developers and deployers to undertake an “Ethical Impact Assessment”.

IEEE in “Ethically Aligned Design” [L1] and UNESCO in “Recommendation on the Ethics of Artificial Intelligence” [L2] (see Figure 1) are referring to key values like “respect, protection and promotion of human rights and fundamental freedoms and human dignity”, “environment and ecosystem flourishing”, “ensuring diversity and inclusiveness”, and “living in peaceful, just and interconnected societies”. Key principles to follow are, among others, “proportionality and do no harm”, “safety & security, fairness and non-discrimination”, “sustainability, right to privacy and data protection”, “human oversight and determination”, “transparency and explainability, responsibility and accountability”, “awareness and literacy”, and “multi-stakeholder and adaptive governance and collaboration”. Applying ethical design in all steps of the software and systems engineering process is thus required nowadays, while different requirements may stem from the different application domains of software engineering (e.g., healthcare) and functions in system context.

The following paragraphs provide a short overview over the various contributions, grouping them according to their specific topics.

Artificial Intelligence ethics and challenges

Rottembourg uses as motivation the Digital Services Act and the Digital Markets Act to discuss relevant regulations in Artificial Intelligence and their consequences for auditing, focusing mainly on the entailed algorithmic challenges (page 8).

Biegelbauer et al. also stress out the importance of adapting the existing workflow of developing ethical Artificial Intelligence applications. In this respect, they are also presenting the role of the AIT AI Ethics Lab created at the AIT Austrian Institute of Technology (page 9).

Leikas et al. also focus on human-centricity and ethics in the development of Artificial Intelligence. They bring to our attention the Ethics Exercise Tool by the Finnish Centre for Artificial Intelligence that entails anticipation, reflexivity of one's own actions, inclusion, and responsiveness (page 11).

Artificial Intelligence and humans

Teixeira et al. present how machines and humans differ when performing classification tasks, and they study this specifically for classifying technological concepts and papers, in order to examine risks and vulnerabilities (page 12).

Héder stresses the need for the software and systems engineer of the 21st century to have knowledge of humanities and social sciences, and presents the Human-Centered AI Masters Programme launched this year in four countries towards this direction (page 14).

Ethical aspects of online data and web

Rossi focuses on the privacy and protection obligations of online data. The author presents the approach followed to gather examples of manipulative designs that can be found in online services and discusses the ethico-legal implications (page 15).

Lafon presents Ethical Web Principles that have been created by the Technical Architecture Group of W3C. The author explains the ethical aspects entailed when building web technologies, applications, and websites, reminding that human rights, dignity, and personal agency also need to be supported (page 17).

Green computing

Noureddine et al. stress out the importance of reducing the energy and carbon footprint of software, and present a number of tools that assist software engineers in measuring and optimising the energy consumption of software systems (page 18).

Radersma discusses code optimisation for minimal energy consumption and gives practical recommendations that software engineers can follow in their everyday activities (page 19).

Dolas et al. address energy in the context of scientific computing ecosystems. To show how they are addressing this issue, they present as case study a strong scaling study on the computational fluid dynamics solver Palabos (page 20).

System design fairness and privacy

Valoggia presents the concept of a risk-based approach to engineering as a means to avoid the pitfalls of engineering to fail by not addressing properly the multidisciplinary characteristics of complex advanced technologies-based systems, especially in the context of Artificial Intelligence and Autonomous Systems (page 22).

Gornet et al. examine aspects that affect the design of a facial authentication system in the context of fairness, and discover aspects that concern data processing, neural networks used, and the training and evaluation of the systems (page 24).

Strobl et al. present an approach for data security by design in the implementation of a privacy-preserving biometric matching system. They describe three possible technical solutions: a reliable system, authentication through a distributed ledger, and use of homomorphic encryption methods (page 25).

Designing for sustainability and pandemics management

Ferrari et al. introduce requirements engineering research for sustainability and present how this can be achieved using interviews from 30 cross-disciplinary experts in the representative domain of digitalisation in forestry, agriculture, and rural areas (page 27).

Rainer et al. present the “ROADS to Health” project that focuses on measures and requirements for addressing cases of pandemics from different views in order to cover various needs. Ethical aspects and human needs are identified as key aspects in this context (page 28).

Note: A related topic “Machine Ethics” was discussed in the “Research and Society” section in ERCIM News 122, July 2020, p. 4 – 11.

Links:

[L1] Ethically Aligned Design, Version 2, IEEE, retrieved Oct. 2022: https://kwz.me/hqD

[L2] Recommendation on the Ethics of Artificial Intelligence, UNESCO, 2021, retrieved Oct 2022: https://kwz.me/hqF

Please contact:

Georgia Kapitsaki
University of Cyprus, Cyprus
gkapi@ucy.ac.cy

Erwin Schoitsch
AIT Austrian Institute of Technology, Austria
Erwin.Schoitsch@ait.ac.at

The Need for AI Regulation and Audit Tools for Bias and Disloyalty Detection

by Benoit Rottembourg (Inria)

The algorithms of large digital platforms massively influence our daily lives. Biases and unfair decisions are increasingly numerous and irritating. The evolving regulatory framework will give rise to new means of auditing, such as black box auditing, which calls for new algorithmic challenges such as those we tackle at Regalia.

The final vote in the European Parliament on July 5th, 2022 on two pieces of legislation, the Digital Services Act (DSA) and the Digital Markets Act (DMA), marks a turning point in the regulation of large digital platforms and their algorithms.

Three main trends have probably led to this evolution of European law: i) The dominant position that certain major digital players, such as Google, Facebook and Amazon, have taken in their markets. They alone represent more than 50% of the global online advertising market (excluding China). ii) At the heart of the value proposition of these platforms are algorithms, often based on artificial intelligence, which allow for both efficient scaling and extreme personalisation of the service. This algorithmic omnipresence makes the practice of regulation authorities both operationally delicate and legally complex. iii) A growing awareness among the general public of the sometimes opaque, even arbitrary or biased nature of the decisions made by these algorithms. The feeling that a form of discrimination or unfair behavior is being carried out by these algorithms is increasingly shared by public opinion. We can think, for example, of the scandal that affected Instagram when it was noted [L1] that images of "curvy" female models in bikinis were more frequently censored than the equivalent images of other models.

Undoubtedly, this evolution of the regulatory framework will have both organisational and technical impacts. The regulatory authorities (European and then national) themselves will see their prerogatives evolve to face this new need for algorithmic compliance. European commissioner Thierry Breton [L2] recently announced that the European Digital Services Act will force tech giants to open the hood of their algorithms so that a committee of experts appointed by the commission can analyse them.

Just as the banking sector experienced in the 2000s and 2010s (with Basel II and Basel III, following the 2008 financial crisis), it is reasonable to think that the major digital players will have to adapt their IT production processes to the new regulatory situation. Audits, for the major algorithms structuring customer interaction, such as recommendation, pricing or moderation algorithms, will have to be carried out by companies in a systematic way. It should be noted that in the text of the DSA the word "audit" appears more than a hundred times [L3]. Even if the technical aspects of these audits are still to be defined by the regulation authorities, platforms should expect to provide greater access to their algorithms and the data that feeds them to accredited experts and academics, without being able to oppose the general terms of use. These steps could go as far as the targeted certification of the algorithms by trusted third parties.

While it is futile to believe that exhaustive transparency of the algorithms used by players at the cutting edge of artificial intelligence and data science can be achieved, it is possible to believe that black box tests, sufficiently sophisticated and piloted by experts, will be able to identify biases or unfair behavior. In any case, this is the direction of a growing amount of research work on the theme of "black box algorithm auditing".

Proving the existence – or absence – of a bias for an algorithm known only as a black box (i.e. by external querying) raises a set of difficult questions ([1],[2]). This is the core challenge of the Regalia project at Inria. We want to highlight here the most constraining characteristics:
- First, the test carried out must be "conclusive", the queries made must be statistically representative and reveal the real activity of the algorithm.
- The queries must not be easily identified by the platform as fictitious behavior, which would allow it to adapt or modify its response.

Figure 1: Algorithms watching algorithms. Image by Gerd Altmann from Pixabay.

- The test must cover a sufficiently large behavioral space, across all possible queries, so that significant prejudice is identified.
- Finally, the test must be frugal so as not to disrupt the platform, and to offer reasonable calculation times for the auditor.

It is therefore understandable that performing a black box audit verifying these properties probably requires advanced algorithmic skills as well as computing resources and human expertise. The Multi-Armed Bandit Problem, a famous mathematical problem of the 1950s and classic benchmark for Reinforcement Learning algorithms, is among the conceptual frameworks for tackling such decision-making problems under uncertainty [3]. Funnily, it is also used by advertising recommendation algorithms of online platforms.

Beyond the necessary engineering and research efforts, critical masses of transdisciplinary personnel will have to be gathered to "lift the hood" of the algorithms of large platforms and offer ammunitions to the regulation authorities.

Links:
[L1] https://kwz.me/hqE
[L2] https://kwz.me/hqK
[L3] https://kwz.me/hqS

References:
[1] E. Le Merrer, R. Pons, G. Trédan, "Algorithmic audits of algorithms, and the law", 2022, arXiv:2203.03711.
[2] B. Ghosh, D. Basu, K. S. Meel, "Justicia: A stochastic sat approach to formally verify fairness" in Proc. of the AAAI Conference on Artificial Intelligence, 2021.
[3] B. Rastegarpanah, K. Gummadi, M Crovella, "Auditing Black-Box Prediction Models for Data Minimization Compliance", NeurIPS 2021.

Please contact:
Benoit Rottembourg, Inria, France
benoit.rottembourg@inria.fr

Ethical AI: Why and How?

by Peter Biegelbauer, Anahid Jalali, Sven Schlarb, and Michela Vignoli (all AIT Austrian Institute of Technology)

Given the numerous opportunities provided by rapidly evolving digital innovations, we need to address and assess the social and political risks that come with naively applying AI algorithms, especially in high-risk sectors. We argue that though the existing guidelines and regulations are a good starting point, we still need to implement effective solutions that can be integrated into the current workflow of developing ethical AI applications. We introduce the idea of AI Ethics Labs as institutionalised "spaces for doubt" providing platforms for a frequent and intensive collaboration between developers and social scientists, thus reducing the potential risks of developed algorithms.

In our digital society, we adopt and integrate information and communication technologies for use at home, work, and in education. The internet offers an unprecedented source for sharing data and collecting information from a broad audience worldwide. AI software provides new means to cope with an increasing amount of data and thereby enables new forms of knowledge production, process optimisation, and decision making. AI solutions have the potential to improve services in a broad range of areas, such as health, agriculture, energy, transportation, retail, manufacturing, and public administration.

However, the latest digitisation wave, and especially AI, raises a number of challenges that need to be tackled. Widespread AI applications such as face recognition, the optimisation of mobility needs, the analysis of employee productivity, and speech recognition are based on the usage of personal data, raising questions regarding privacy, data protection, and ethics. Decision-supporting algorithmic systems may be and often are based on datasets containing biases, from which algorithms learn, and may also include explicitly discriminatory state-

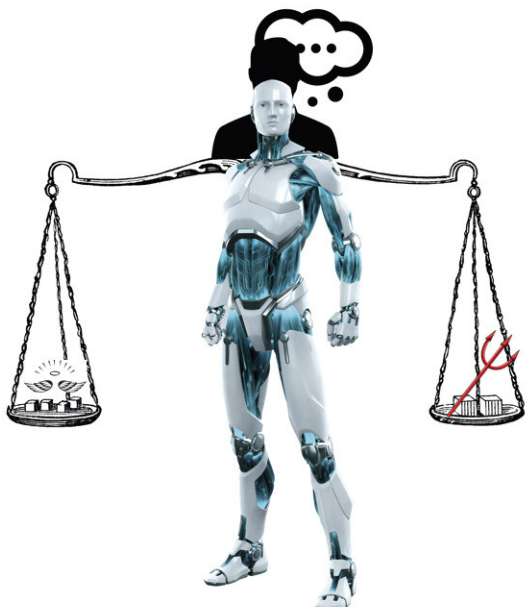


Figure 1: Ethical guidelines to develop an AI ensures a fair system.

ments affecting predictions of AI applications. How serious the problem is becomes evident when we look at recent examples of political scandals that resulted from ill-used, biased algorithmic solutions using personal data that had led to financial and legal harm to innocent citizens (e.g., Robodebt in Australia, the childcare benefit scandal in the Netherlands) [L1][L2]. In Austria, the planned introduction of a job-market opportunities assistance system (AMAS) was heavily criticised due to its inherent biases and other implications [1].

We can see that our society, economy, and industries are being heavily impacted by AI technologies. However, the inherent risks and implications are often underestimated or neglected. In order to sustain an inclusive, secure, and socially just digital society that benefits from AI technologies, we need to establish clear regulations and guidance on how to develop and implement ethical AI solutions.

Recently, several ethical principles, guidelines and regulations have been proposed by national and international governmental bodies, with the EU being at the forefront of such activities. The EU digital package, containing literally dozens of regulations under discussion in 2022, includes the AI Act [L3], which focuses on AI and its impact on society.

In addition, professional associations such as the Institute of Electrical and Electronics Engineers (IEEE), the Association for Computing Machinery (ACM) and the North American Chapter of the Association for Computational Linguistics (NAACL) have repeatedly been engaged with AI-related norm-setting.

Recent research, however, has shown that ethics guidelines have only a very limited impact on the practice of software engineering [2][3]. The guidelines lack more practice-oriented guidance for addressing the complex impact of AI and perpetuate the disciplinary divide between social scientists and ethicists who produce the guidelines, and technical engineers who write the software. Common standards for creating more ethical AI solutions should integrate recommendations for technical solutions (e.g., bias identification), “ethics by design” principles, and self-assessment tools (e.g., ethics committees). Another issue is the availability of resources for thinking about ethics in context of the software development work – a factor regularly missing in the actual work of software engineers.

We see a need for institutionalised “spaces for doubt” [L4], which may be provided by inter- and transdisciplinary groups. Regular exchange between experts from different technical and non-technical disciplines as well as stakeholders is needed to build common standards and trust. At the AIT Austrian Institute of Technology, we created the AIT AI Ethics Lab [L5] in 2020, which since then has served as a platform for regular meetings between software engineers and social scientists. Many of these interactions took place between experts from AIT itself, serving as a forum of learning together about the ethical challenges of AI development. More recently, activities have involved also actors from other organisations and sectors, leading to cooperation with various actors such as the Austrian Federal Academy of Public Administration (VAB), the Austrian Federal Ministry for Civil Service (BMKOeS), and TAFTIE, the umbrella organisation of European innovation agencies.

The risks associated with the incorrect use of AI technologies are increasingly being perceived as endangering the acceptance of AI. Given their increasing importance, the ability to be compliant with ethical guidelines will turn into a competitive advantage – even more so if compliance with binding regulations will be demanded by authorities and companies in future. The model of AI ethics labs allows for an ongoing self-reflection regarding software development, and it provides the possibility for software engineers, social scientists, management and clients to engage in a productive exchange. An adequately financed ethics lab can contribute to address the above-outlined challenges of AI and propel the AI ethics debate from theory to practice.

Links:
[L1] https://kwz.me/hq4
[L2] https://kwz.me/hq0
[L3] https://kwz.me/hq1
[L4] https://kwz.me/hq2
[L5] https://cochangeproject.eu/labs/AIT

References:
[1] D. Allhutter, et al., “Der AMS Algorithmus - Eine Soziotechnische Analyse des Arbeitsmarktchancen-Assistenz-Systems”, AMAS, p. 120, Wien, 2020. doi:/10.1553/ITA-pb-2020-02.
[2] P. Biegelbauer, et al., “Engineering Trust in AI: the impact of debates on AI regulation on the work of software developers”, in 20th STS Conference Graz 2022, Critical Issues in Science, Technology and Society Studies, 2022.
[3] T. Hagendorff, “The Ethics of AI Ethics: An Evaluation of Guidelines”, in Minds and Machines 30(1): 99-120, 2020. doi:10.1007/s11023-020-09517-8.

Please contact:
Peter Biegelbauer, AIT Austrian Institute of Technology, peter.biegelbauer@ait.ac.at

Anahid Jalali, AIT Austrian Institute of Technology, anahid.jalali@ait.ac.at

A Systematic Exercise Tool Helps Researchers Ponder the Ethical Implications of AI

by Jaana Leikas (VTT Technical Research Centre of Finland Ltd) and Patrik Floréen (University of Helsinki, Finnish Centre for Artificial Intelligence)

In addition to many clear benefits, AI technology can also bring risks related to the misuse of technology and for example increased inequalities. There is thus an increased demand to apply the principles of human-centricity and ethics in the development of AI. Members of the Ethics Advisory Board of the Finnish Centre for Artificial Intelligence (FCAI) have developed an Ethics Exercise Tool to help AI researchers consider and discuss the ethical implications of their research and strengthen ethical understanding within their team.

Although AI is a generic technology, ethics of AI can be understood as a contextual phenomenon. Understanding the context means that the dynamic effects of AI should also be understood. For example, contextual relevance is indicated by the fact that biased data can be almost irrelevant in one context (if it is not used for any activity) and extremely relevant in another.

Problems related to social contexts and human behaviour are central in AI, placing emphasis on value questions such as the status of citizens, democracy, and fairness. These issues are often multidimensional, even ambiguous, and require an analytical approach and new tools to operationalise and integrate them into the context of AI.

Many public, private and non-governmental organisations and expert groups have provided views and guidelines for the design of ethically and socially acceptable AI. For example, The European Commission's independent, high-level expert group on artificial intelligence (AI HLEG) states in Ethics Guidelines for Trustworthy AI [L1] as the ethical purpose of the document to "ensure respect for fundamental rights, principles and values when developing, deploying and using AI". One of the most recent ones, the UNESCO Recommendation [L2], starts from the premise that there are ethical issues at every stage of the life cycle of AI systems. This holistic view of the ethical development of AI is reflected in the fact that the ethical consideration is embedded in the practical activities of people by including not only AI systems and their development, but also the issues related to their use. The recommendation refers to all actors involved in at least one stage of the AI system life cycle as "AI actors". This includes both natural and legal persons, such as researchers, programmers, engineers, data scientists, and end-users, as well as businesses, universities, public authorities, and private entities.

Proactive assessment of the impact of AI is necessary when discussing ethics. This requires insight into both short-term and long-term challenges. For this purpose, the Finnish Centre for Artificial Intelligence (FCAI) has launched an Ethics Exercise Tool [L3] to help AI researchers identify, explicate, and generally work with ethical issues related to their work, as well as understand the ethical and social implications of their research. The tool developed by FCAI ethics experts follows the principles of responsible research and innovation. The tool is a frame for systematic thinking of futures that are and are not desired. It helps in avoiding pitfalls and in guiding research toward the greater common good. The perspective from which it is examined varies from individual projects to entire research programmes, and it is to be used iteratively in the planning, implementation, and communication of research. It comprises various thematic questions that researchers should consider in their teams in the different phases of a project.

The first, general goal of the tool is to link the results of the research to a broader societal context through ethical consideration. The second immediate goal of the tool is to help researchers identify and resolve ethical and societal problems related to their own work; consideration of the effects is also required by an increasing number of publication forums and funding instruments of research.

Stimulating debate is important

The tool is based on the assumption that ethical consideration cannot be outsourced, because of its dependence on context: answering ethical questions about the problem in question is the responsibility of researchers working with the problem. Therefore, researchers need to awaken to ethics and try to sufficiently understand the field from which ethical questions emerge. A basic understanding of ethics is starting to be essential also for students of artificial intelligence.

The idea of the FCAI Ethics Exercise Tool is based on four complementary points of view. Anticipation, involving careful consideration of the sought-after and accidental consequences of the research and innovation activities. Reflexivity of one's

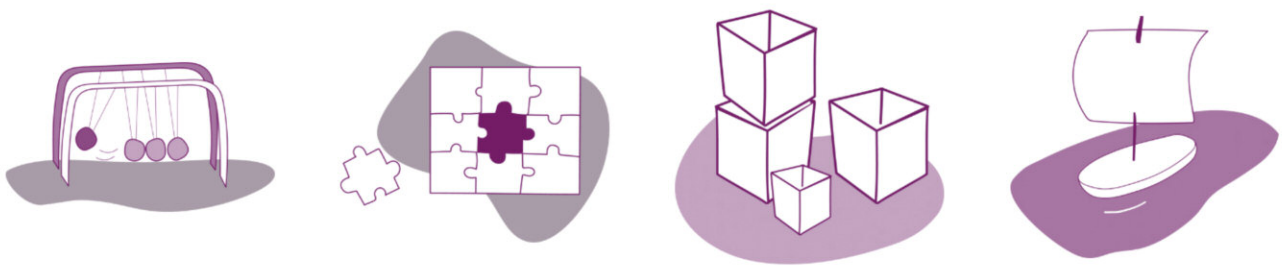


Figure 1: The FCAI Ethics Exercise Tool includes reflections on anticipation, reflexivity, inclusion, and responsiveness.

own actions, in which possible assumptions and commitments are considered, which might affect the research work. Inclusion, in which the relevant stakeholders are invited to engage in dialogue on the desired and unwanted consequences of the research. Predisposition and responsiveness: seeking the right direction for the research in accordance with the visions and values that have emerged through the processes that have been highlighted.

Using the tool may aid in fostering ethics at critical phases in the project life cycle. It should be used iteratively, where ethical issues are systematically discussed before initiating and after the project. Reflection on ethics should thus include reflection on the ethical values and choices in respect to design decisions, as well as the possible impacts of research outcomes on a societal level.

Like understanding the ethics of AI, developing an ethics exercise tool also requires a continuous dialogue between researchers. FCAI therefore wants to iteratively develop and improve the tool and thus continuously seeks feedback from researchers on its use.

Links:
[L1] https://kwz.me/hq3
[L2] https://en.unesco.org/artificial-intelligence/ethics
[L3] https://fcai.fi/ethics-exercise

Reference:
[1] A. Sigfrids, et al., “How should public administrations support ethical and sustainable development and use of Artificial Intelligence? A systematic review of proposals for developing governance of AI”, Frontiers in Artificial Intelligence, 2022.
https://doi.org/10.3389/fhumd.2022.858108

Please contact:
Jaana Leikas
VTT Technical Research Centre of Finland Ltd., Finland
jaana.leikas@vtt.fi

Humans Versus Machines: The Perspective of Two Different Approaches in Classification for Ethical Design

by Sónia Teixeira, José Rodrigues (Faculty of Engineering of University of Porto and INESC TEC), Bruno Veloso (Portucalense University and INESC TEC) and João Gama (Faculty of Economics of University of Porto and INESC TEC)

This Portuguese project compares the classification of AI risks and vulnerabilities performed by humans and performed by the computing algorithms.

There has been an increase in the use of Artificial Intelligence (AI) technologies, whether at home, in public spaces, in social organisations, or services. The growing adoption of these systems, in particular the data-driven decision models, has called attention to risks arising from the use of technology, from which ethical problems may emerge. However, despite all the research, the definition of concepts and nature inherent to those risks/vulnerabilities is not consensual. Therefore, categorising the vulnerability type of those systems will facilitate an ethical design.

In this project, we compare the classification of AI risks/vulnerabilities performed with two different approaches: the classification performed by humans and the classification performed by the machine. This comparison puts into perspective the similarities and differences between the classification in both approaches. The main goal of this work is to understand which types of AI risks/vulnerabilities are ethical and which are technological, as well as to identify the differences between human versus machine classification. This initial step may bring insights to analyse the bias or other challenges related to software development in the future and the human aspects involved in the decision to be considered in designing AI-based systems.

Considering the published journals that mention the risks, vulnerabilities, and challenges of AI, we assume technological risks as those focused on the technical issues, and ethical risks as those that arise in the outcome, focusing on the non-technical issues.

Our approach [1] considers a literature review and a selection of articles from three experts from different areas in the first stage. The second stage involved carrying out a survey with questions for classification by humans, in which we included the risk concepts identified in the first stage. Finally, the third stage involved using an algorithm recognised in the literature as a good baseline for text classification using machine classification. In this step, we used the papers selected from the literature review, from which we extracted the risk concepts used in the second stage for the algorithm classification. Once having the results of the classification by humans and machines,

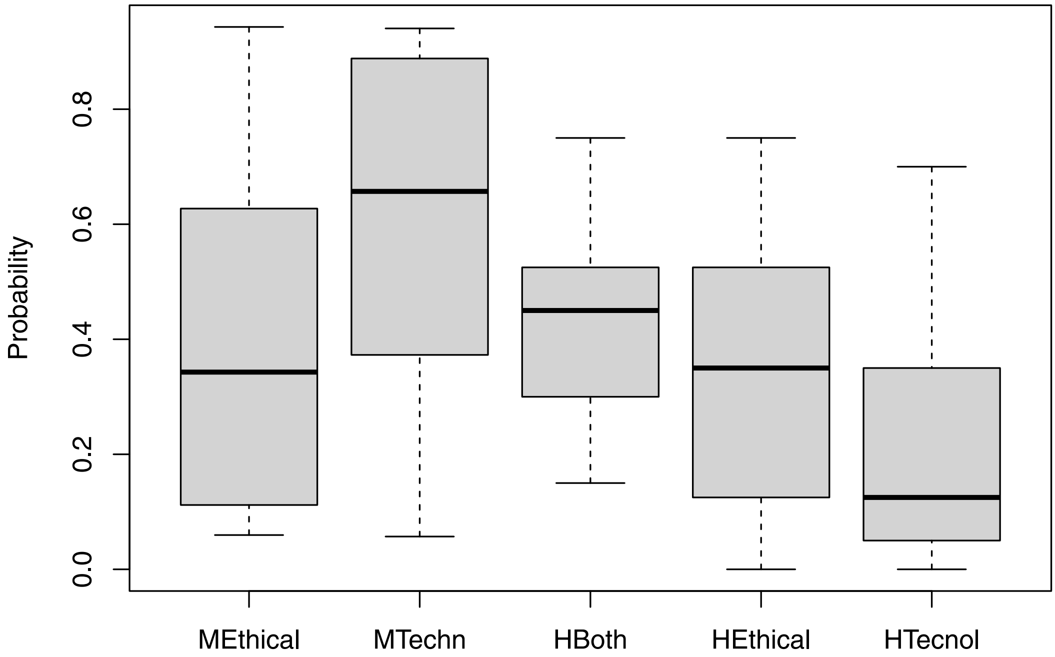


Figure 1: Distribution of the Machine(M) and Humans(H) classifications.

we resort to i) Descriptive Data Analysis, ii) Multiple Factor Analysis (MFA), and iii) Clustering for the analysis of these classifications.

In this project, we observe that machine and human classification distribution (Figure 1) seem to be different in classifying technological concepts and papers. However, in the case of classifying concepts and papers as ethical, the median of human and machine classifications is very close. The results show that we have three clusters. Cluster one comprises six of the ten risks/vulnerabilities in which humans and machines did not reveal consensus. Cluster one corresponds to the concepts: bias, interpretability, protection, explainability, semantic, opacity, completeness, accuracy, data quality, and reliability. Cluster two consists of eight risk concepts in which humans and machines agreed in their classification, a classification of a technological nature. This cluster includes concepts such as extinction, transparency, fairness, manipulation, safety, and security. Finally, cluster three includes risk concepts essentially classified as ethical. This is the case for six risks, which humans and machines rated as ethical. Vulnerabilities such as moral, power, diluting rights, responsibility, systemic, liability, accountability, and data protection belong to cluster three. In the case of humans, the classification of vulnerability concepts is carried out at a more abstract level, and in the case of the machine, it is carried out at a more contextual level. I.e., even with different levels of detail for the classification, the classification of vulnerabilities is in agreement in most cases.

In the future, we intend to deepen our research to understand better the sensitivity of human classification, the machine model, and the resulting bias.

This project is part of a PhD thesis in Engineering and Public Policy, started in 2019, at the Faculty of Engineering of the University of Porto, with the host of the Laboratory of Artificial Intelligence and Decision Support from INESC TEC. The research reported in this work was partially supported by the European Commission funded project “Humane AI: Toward AI Systems That Augment and Empower Humans by Understanding Us, our Society and the World Around Us” (grant \#820437). The support is gratefully acknowledged.

Reference:

[1] S. Teixeira, et al., “Ethical and Technological AI Risks Classification: A Human vs Machine Approach”, in ECML PKDD 2022 Workshops.

Please contact:

Sónia Teixeira, INESC TEC, Portugal
sonia.c.teixeira@inesctec.pt

The Engineers of the 21st Century

by Mihály Héder (ELKH SZTAKI and Budapest University of Technology and Economics)

This paper reports on a novel curriculum, the Human-Centered AI Masters Programme (HCAIM). HCAIM is launched this year in four countries. It incorporates humanities, social sciences and law into the education of Artificial Intelligence in an unprecedented amount and depth.

The oncoming new wave of ethics recommendations and regulation for technology will present an unprecedented intellectual challenge to software and systems developer teams worldwide, but particularly in Europe where consumer protection and the social control of technology is especially stringent. This challenge will require surmounting the cultural and sometimes even cognitive barriers between engineering and the humanities.

The challenge comes in the form of interdisciplinary language and content in normative documents and in the resulting need for professionals who can read and interpret such texts. For instance, the IEEE Ethically Aligned Design document starts out with a presentation of the Aristotelian concept eudaimonia; the IEEE P7000 standard series draft includes several categories from moral theory like deontological and consequentialist ethics. The upcoming AI regulation of the European Union heavily relies on the concepts of transparency and trustworthiness, an epistemic concept and a virtue; the EU High-Level Expert Group's Ethics Guidelines for Trustworthy AI borrows a page from contemporary social science and replaces the profit motive of the corporation with chasing a wellbeing metric instead. And the list continues with the concepts of power asymmetry, social fairness, biases and so on [1] in the OECD, UNESCO, etc. regulation, the discussion sometimes going as far as the metaphysics of machines [2], which was unthinkable before in the subculture that engineers create for themselves, based on mathematics, physics and a strong sense of down-to-earth systems thinking.

The normative documents of engineering activities were never this interwoven with philosophical and political language. Yet, all of these elements are the result of public consultations – the self-regulation of professional societies and the influence of elected representatives. In other words, we can take these documents as expressions of society's need for social control of technology [3]. If so, this need may reflect the discovery of the hidden power of engineers, especially those who develop software (AI included). The special property of software is that it is a machine that can be manufactured with almost zero marginal cost once the design is finalised [3, p127] – in fact it is not even called manufacturing, just copying.

As a result, software is prone to create technological lock-in situations, because the incentives are extremely high for the reuse of pre-existing software instead of writing new if the licensing and modularity is also right (a great example is the dominance of the Linux kernel). Software-as-a Service (SaaS) enhances this effect further, by minimising the marginal cost of serving one more request through economics of scale. Our contemporary software tools for productivity, collaboration, AI and even science reflects this: unless there are prohibiting factors (i.e., software needs to run on-board) we tend to use SaaS. The ethical recommendation and regulation tsunami is the reaction to this extreme lock-in potential.

To tackle these challenges, the software and systems engineer of the 21st century needs to be a person with deep knowledge about the humanities and social sciences. Efforts need to be undertaken to properly educate these engineers, but this requires a common language and understanding first among the educators. The Human-Centered AI Masters [L1] is such an effort. The curriculum is created by an interdisciplinary team involving four universities and five companies and is taught in (in alphabetical order) Budapest, Dublin, Napoli and Utrecht from 2022. While the curriculum is AI-specific, its structure is

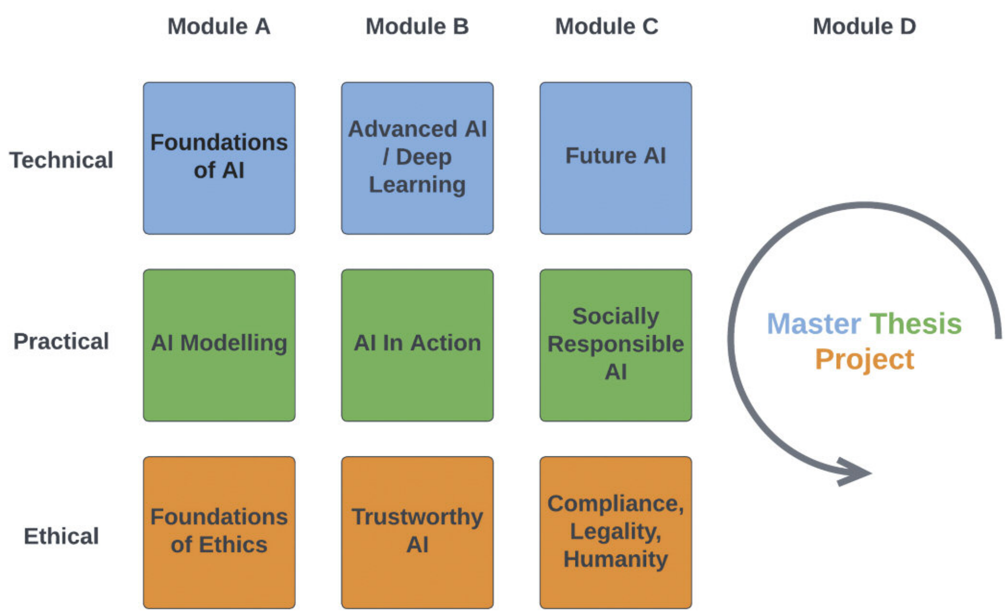


Figure 1: The overview of the HCAIM program: after interdisciplinary studies the students write a synthetising master's thesis.

not far from what ethics for any kind of software needs. Unlike other engineering curriculums, it includes a significant amount – around 25% – of outright humanities, broadly construed, including legal, economic and social aspects.

Moreover, the more technology-specific subjects and all case studies also are aware and make use of foundations of ethics, social sciences and law. The development of the curriculum required ground-breaking work on the incorporation of topics such as social responsibility, transparency and explainability; value-based design; and bringing social fairness into these rethought technical subjects.

Work done in the field of engineering education for the previous wave of regulation, mostly about privacy and data governance, is also incorporated. As a result, privacy-preserving machine learning, algorithmic justice and the prevention of bias may be learned, along with forward-looking topics such as future AI topics including singularity, robot rights movements, and human-machine biology are also discussed.

The Human-Centered AI Masters programme was Co-Financed by the Connecting Europe Facility of the European Union Under Grant №CEF-TC-2020-1 Digital Skills 2020-EU-IA-0068.

Link:
[L1] https://humancentered-ai.eu/

References:
[1] M. Héder, “A criticism of AI ethics guidelines”, in Információs Társadalom XX, no. 4 (2020): 57–73. https://dx.doi.org/10.22503/inftars.XX.2020.4.5
[2] M. Héder, “The epistemic opacity of autonomous systems and the ethical consequences”, in AI & Society (2020). https://doi.org/10.1007/s00146-020-01024-9
[3] M. Héder, “AI and the resurrection of Technological Determinism”, in Információs Társadalom XXI, no. 2 (2021): 119–130. https://dx.doi.org/10.22503/inftars.XXI.2021.2.8

Please contact:
Mihály Héder, ELKH SZTAKI and Budapest University of Technology and Economics
mihaly.heder@sztaki.hu

Finding a Way Out the Ethico-Legal Maze of Social Media Data Scraping

by Arianna Rossi (SnT, University of Luxembourg)

People know that once they publish something online, such content becomes "publicly available" and can be downloaded and re-used by others, for example, researchers and data scientists. The reality is far more complicated. And for us, finding a way to comply with data protection obligations and to respect the tenets of research ethics became an exploration of a largely uncharted territory.

Within an interdisciplinary project (Dark Patterns Online "DECEPTICON" [L1]) carried out at the SnT, University of Luxembourg, together with the Human-Computer Interaction group and the Luxembourg Institute of Science and Technology, we are gathering examples about the many manipulative designs that populate online services (i.e., dark patterns, see Figure 1) and are publicly condemned on Twitter and Reddit. Our aim is to build a labelled dataset of such pervasive practices by using crowdsourced knowledge and possibly develop supervised machine learning models to flag dark patterns at scale.

Initially, we were convinced that we only needed to address a few data protection concerns, which seemed totally feasible. However, we found out that there is a plethora of legal obligations to comply with and additional research ethics principles to be considered. Finding creative answers to such issues was a long, tiresome, albeit formative experience that we briefly share in these pages, with the conviction that it can be of help to other academic and industrial researchers who collect and analyse internet data.

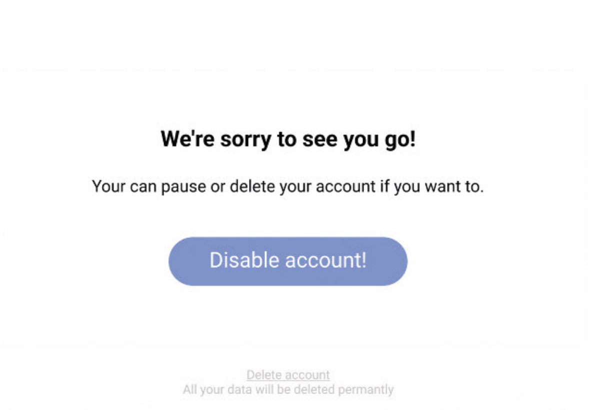


Figure 1: When a user tries to delete their account on this platform, they are confronted with two options: the "disable account" button is very visible, while the "delete account" option is greyed out and hidden at the bottom of the page. As we showed in our study [1], many people do not even notice the second option. Moreover, the consequences of disabling as opposed to deleting the account are unclear: for instance, will the personal data linked to the account be erased? The example has been freed to any reference to brands and companies.

First, we need to ask: what may count as personal information on social media? The answer is: Potentially, almost everything. Not only the username and pictures can reveal a users' identity, but also metadata like the timestamp, the location and URLs contained in tweets and online posts can trace back to specific people. Moreover, a simple search online of a tweet or part of it can easily lead to its authors and all the associated information. It also means that merely removing usernames from a dataset does not equal anonymising data, as data scientists often argue in their papers. Consequently, we fall within the realm of the General Data Protection Regulation and thus we should observe many obligations about the transparent, lawful and fair processing of personal data.

To enhance data confidentiality, for example, many security measures must be implemented, like encrypting the dataset and only using encrypted channels to transfer it to a private Git repository, which must be subject to strong authorisation and access-control measures. We also pseudonymised the data, i.e., we masked or generalised data like personal names, locations and timestamps, and implemented a re-identification function that can re-establish the original data at will, for example to retrieve the authors of social media posts and allow them to opt out from our study. We are now examining more advanced and secure techniques for pseudonymisation.

We also experimented innovative ways to be transparent: we published a privacy policy that followed best practices of legal design, for instance a conversational tone that clarifies what are the responsibilities of the researchers as opposed to the rights of the social media users, and we emphasised the main information to allow skim reading. Since such a privacy notice is tucked away on our project's website, the possibility that a Twitter user stumbles upon it is extremely weak, so we set an automated tweet that once a month alerts about our invisible data collection and gives practical instructions on how to opt out of it.

Apart from these and other data protection measures, we also had to embed research ethics into our activities. Contrary to what many scientists believe, research on internet data counts as research on human subjects, and must therefore offer the same level of safeguards. However, data scraping mostly happens without the knowledge of the research participants who thus don't have the option to freely decide whether to take part in the study or not. Since we extract content that has been disclosed in a certain environment, sometimes within a closed community, we need to foresee the possible consequences of extrapolating, reusing and disclosing such information in a different context. Given that seeking the informed consent of thousands of people may be impossible, we tried to be very transparent about our activities and gave the possibility to opt out of the study. This is why we pseudonymised the data, so that we could exclude certain parts of them on demand. Additional details, including other issues like how to treat minors' data, how to address cyber risks and how to attend to data quality in such specific settings are described in [2].

Reflecting on our time-consuming experience of producing innovative multidisciplinary solutions, we asked: given the time pressure of the academic world and that unethical and illegal behaviour is only rarely sanctioned, what kind of incentives could encourage internet data researchers to go through the same pain as us? Training, even when based on immersive, entertaining experiences, is necessary, but not sufficient: the intention–behaviour gap remains unsolved and when procedures are too complex, human beings find workarounds to what they perceive as obstacles.

Thus, the second order of solutions should make compliance and ethical decision-making less burdensome. Practical guidance drafted in laymen terms, as well as best practices and toolkits tailored to certain technologies should be created and proactively brought to the attention of researchers. Moreover, we need usable off-the-shelf solutions that simplify and expedite academic compliance tasks. We are now working on an open-source Python package for social media data pseudonymisation [3].

We start to see the light at the end of this ethics-legal maze and hope that our Ariadne's thread will guide other legality-attentive researchers [L2] out of it.

Links:
[L1] https://kwz.me/hq8
[L2] https://www.legalityattentivedatascientists.eu/

References:

[1] K. Bongard-Blanchy, et al., “I am Definitely Manipulated, Even When I am Aware of it. It’s Ridiculous! - Dark Patterns from the End-User Perspective”, in Designing Interactive Systems Conference 2021 (pp. 763-776). 2021.
[2] A. Rossi, A. Kumari, G. Lenzini, “Unwinding a Legal and Ethical Ariadne’s Thread out of the Twitter’s Scraping Maze”, in Privacy Symposium 2022—Data Protection Law International Convergence and Compliance with Innovative Technologies (DPLICIT), S. Ziegler, A. Quesada Rodriguez and S. Schiffner, Springer Nature (in press).
[3] A. Rossi, et al., “Challenges of protecting confidentiality in social media data and their ethical import”, in 2022 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW), pp. 554-561, IEEE, 2022.

Please contact:

Arianna Rossi, SnT, University of Luxembourg, Luxembourg
Arianna.rossi@uni.lu

W3C Technical Architecture Group Ethical Web Principles

by Yves Lafon (W3C)

The Technical Architecture Group (TAG) is a special working group within the World Wide Web Consortium (W3C). Its mission is to document and build consensus around principles of web architecture to help ensure that the web makes sense as a platform and that the overall design is coherent.

To achieve this, the TAG [L1] provides reviews of all specifications produced at W3C, and some outside of W3C. During those reviews, patterns emerged, like common issues, or unintended harmful consequences.

As such, the TAG produced a series of documents, called principles, to address those issues early-on. They are part of the TAG's publications known as Findings (short documents that cover a specific issue in the web architecture). Most principles are purely technical, but the basic principles started as general ethical guidelines, like "Put user needs first (Priority of Constituencies)", "It should be safe to visit a web page" or "Leave the web better than you found it". A wider list of those principles led to the creation of the Ethical Web Principles.

The initial observations from the TAG became of interest for the whole W3C membership, leading to the publication of those documents in the Statement track, meaning that the goal is to get the consensus of the W3C membership on those documents. This document and the set of principles included are not only intended for specification developers, but also for website authors, tools designers, etc.

Principles

The web should empower an equitable, informed and interconnected society. It has been, and should continue to be, designed to enable communication and knowledge-sharing for everyone. In order for the web to continue to be beneficial to society, we need to consider the ethical implications of our work when we build web technologies, applications, and sites.

The web is made up of several technologies and technical standards. HTML, CSS and JavaScript are often thought of as the web's core set of technologies but there are a raft of other technologies, standards, languages, and APIs that come together to form the "web platform". One of the web platform's differentiators has always been a strong ethical framework; for example, an emphasis on internationalisation, accessibility, privacy, and security.

Web technologies are also offered royalty free to enable open source implementation. These are often cited as some of the strengths of the web. Despite this, in the 30 years since development of the web began, it has become clear that the web platform can often be used in ways that subvert its original mission, or even be used to cause harm.

The architecture of the web is designed with the notion of different classes of application that retrieve and process content, and represent the needs of the application's users. This includes web browsers, web-hosted applications such as search engines, and software that acts on web resources. This lends itself well towards empowering people by allowing them to choose the browser, search engine, or other application that best meets their needs (for example, with strong privacy protections).

The web should also support human rights, dignity, and personal agency. We need to put internationally recognised human rights at the core of the web platform. We can reinforce this approach by promoting ethical thinking across the web industry.

The principles in the W3C TAG Ethical Web Principles document [L2] are deliberately unordered, and many are interconnected with each other. They are intended to be viewed holistically, rather than each one in isolation. While all of the principles together aim to provide pillars that collectively support a web that is beneficial for society, there are cases where the effects of upholding one principle may diminish the efficacy of another principle. Thus in applying these principles, there are benefits and trade-offs that may need to be carefully balanced. When proposals that support particular principles appear to be in conflict with other principles, it is important to consider the context in which the technology is being applied, the expected audience(s) for the technology, who the technology benefits and who it may disadvantage, and any power dynamics involved (see also the priority of constituencies).

This document is still being discussed actively and many changes and clarifications are expected in the near future. Feedback is welcome in the related GitHub repository [L3].

W3C TAG Ethical Web principles

- There is one web
- The web should not cause harm to society
- The web must support healthy community and debate
- The web is for all people
- Security and privacy are essential
- The web must enable freedom of expression
- The web must make it possible for people to verify the information they see
- The web must enhance individuals' control and power
- The web must be an environmentally sustainable platform
- The web is transparent
- The web is multi-browser, multi-OS, and multi-device
- People should be able to render web content as they want.

Links:
[L1] https://tag.w3.org/
[L2] https://www.w3.org/TR/ethical-web-principles/
[L3] https://github.com/w3ctag/ethical-web-principles/

Please contact:
Yves Lafon, W3C
ylafon@w3.org

Software Energy Efficiency: New Tools for Developers

by Adel Noureddine and Olivier Le Goaër (Université de Pau et des Pays de l'Adour)

Reducing the energy and carbon footprint of software is a major concern for practitioners and researchers today. But what tools does a student in computer science or a professional developer have at their disposal to improve the energy efficiency of their software? At our LIUPPA laboratory, we create and develop tools to help software developers understand and improve the energy efficiency of their software.

With the rise of the ecological question and sustainable development in the economical and political agenda worldwide, the information technology (IT) sector is receiving increasing attention. Indeed, its ecological impact (energy consumption, induced CO2 emissions and resource exhaustion) cannot be ignored. Among the different layers in modern systems, the role of software is paramount in reducing the energy and carbon footprint of IT industries.

As stated in recent surveys, developers need tools to help them get a better understanding of the energy consumption of software systems at large and the root causes behind it, and then to find efficient ways to improve it. This implies providing tools to measure power and energy consumption, but also tools to make the link between energy measurements and source code in order to improve the software quality from an ecological perspective.

When tackling the energy question, the reality of modern software development involves targeting different platforms (IoT, server, mobile devices), at different times (runtime, design-time), and at different granularity levels (system, software, lines of code).

In our laboratory, we design and build multiple tools aimed at filling the gap in measuring and optimising energy consumption of software systems. We describe, below, our main four tools that are aimed at covering multiple platforms, operating systems, and stages of software lifecycles:

- PowDroid [L1][1] is a command-line tool collecting system-wide energy-related metrics from any Android device plugged through USB to a desktop PC. Metrics include status of energy-hungry components over time (Screen, GPS, Wifi, etc.), along with evolution of physical measures (voltage, charge, intensity, power). It gives real energy consumption of a hardware device. If an app is tested in good isolation (i.e., without running other apps or services), PowDroid can therefore give an estimation of the energy consumption of an app itself. The tool is straightforward to use and generates a detailed and continuous CSV file of the energy consumption of the Android phone.
- PowerJoular [L2][2] is a multi-platform power-monitoring tool that can monitor, in real time, the power consumption of hardware components (CPU, GPU) and software (at the process level). It currently supports Intel and AMD CPUs on GNU/Linux, Nvidia GPUs, and Raspberry Pi's ARM CPUs. PowerJoular uses RAPL for Intel/AMD CPU energy, and regression models we build for ARM CPU energy. The tool can run automatically with a systemd service, and write export data for later analysis. Therefore, it can also be used by system administrators to monitor a fleet of devices (such as devices deployed in an industrial setup, or multiple servers in a data centre). PowerJoular is written in Ada and has low overhead for runtime monitoring.
- JoularJX [L3][2] is a Java-based agent that provides real-time power monitoring of methods in Java applications. The tool supports Linux (using RAPL) and Windows (using Intel's API), and Java 11+. JoularJX exports its power readings in a CSV file on runtime (a file is created and overwritten every second), and therefore allows real-time monitoring of power fluctuations of each method in a Java application. It can also provide total energy readings at the program's end. JoularJX provides valuable input for software developers to diagnose and improve the energy efficiency of their software, with fine-grained and real-time power monitoring that allows tracing an energy profile through time for each method of an application.
- ecoCode [L4][3] is a sonarQube plugin that extends clean code with green code. The most advanced component so far targets native Android projects written in Java, enabling the static detection of 40 energy code smells on any codebase. The android-specific smells catalogue originated from existing research literature, mining of the Android API reference and interviews of senior mobile app developers. It also comes with a customised UI for a new user experience on that topic. New components are already underway for this growing open source project in order to target further technologies (pure java or python programs, iOS projects, etc.).

Our tools series covers a wide spectrum of energy and power tools aimed at helping software developers and administrators monitor and optimise devices and software systems, and to aid developers in designing lower carbon software. We hope our tools help the research community in conducting more empirical software research around green IT, and help practitioners and software developers on both legacy and new systems.

Links:
[L1] https://gitlab.com/powdroid/powdroid-cli
[L2] https://www.noureddine.org/research/joular/powerjoular
[L3] https://www.noureddine.org/research/joular/joularjx
[L4] https://kwz.me/hqy

References:
[1] F. Bouaffar, O. Le Goaer, A. Noureddine, "PowDroid: Energy Profiling of Android Applications", in Proc. of SUSTAINSE/ASE'21, 2021.
[2] A. Noureddine, "PowerJoular and JoularJX: Multi-Platform Software Power Monitoring Tools", in Proc. of IE2022, 2022.
[3] O. Le Goaer, J. Hertout, "ecoCode: a SonarQube Plugin to Remove Energy Smells from Android Projects", in Proc. of ASE 2022, 2022.

Please contact:
Adel Noureddine, Université de Pau et des Pays de l'Adour, France, adel.noureddine@univ-pau.fr

Olivier Le Goaër, Université de Pau et des Pays de l'Adour, France, olivier.legoaer@univ-pau.fr

Green Coding: Reduce Your Carbon Footprint

by Reinder Radersma (NWO-I Digital Competence Center)

Green coding – the practice of optimising code for minimal energy consumption – is a direct way for software engineers to reduce the carbon footprint of their work. This article reflects on current trends in energy efficient computing and gives six recommendations for software engineers to implement in their daily practices.

The Dutch Research Council operates nine national research institutes in the Netherlands. To support their researchers with open and reproducible science, a Digital Competence Center (NWO-I DCC) has been set up, hosted by CWI. NWO-I DCC commits to software being open and available to others, but this comes with the obligation to also address any negative consequences, such as a larger carbon footprint. NWO-I DCC therefore promotes green coding to minimise the environmental impact of research. This article is a summary of insights gathered during a mini-symposium on green coding organised in July 2022.

Our current society is unimaginable without computation, and the need for computing power will only increase. Over the next two decades computation capacity is predicted to multiply a million-fold and consequently its energy consumption will double every three years. Given the current energy and climate crisis, efforts to reduce energy usage and therefore the carbon footprint of computing are imperative.

High performance computing

Energy takes up a substantial part of the maintenance costs of High Performance Computing (HPC) facilities. So apart from environmental considerations, financial incentives have promoted energy efficiency. For instance, processors have become more energy efficient by increasing the number of GFLOPs (a measure of computer performance) per watt. However, this efficiency gain is not on par with the increasing demand for computation power. Other initiatives to reduce the carbon footprint of HPC facilities are, for instance, the use of heat waste for heating buildings.

Not only hardware but also software is used to lower impact. Energy management software is used to tune HPC clusters for lower energy usage. By reducing the clock speed of Central Processing Units (CPUs) a speed reduction of a few percent can reduce total energy usage also by a few percent. By changing the standard settings to more energy efficient values and giving users access to energy management software such as EAR, the carbon footprint can be further reduced [L1].

Local computing

For local computing (such as PCs laptops and local servers) similar trends are visible. PC, laptop and server processors benefit from the same developments as HPC clusters. Particularly for laptops, efficient processors have been developed to reduce (battery) weight while increasing functional time off the energy grid. Measuring energy efficiency is trickier though. This can be done with wattmeters, but alternatively there are CPUs that can measure their own energy usage (albeit ignoring energy usage by memory, etc.) or processor-specific estimates.

The Dutch Research Council (NWO) promotes green coding to minimise the environmental impact of research.

Recommendations for software engineers

When developing and running software, there are many choices that will affect the energy usage of your software. Here are six recommendations for lowering the carbon footprint:

1. Choose a green language
   Some programming languages are more energy efficient than others; it depends on the number of operations that underly commands. Compiled languages (e.g., C, C++, Fortran, Ada) are typically more energy efficient than interpreted languages (e.g., Python, Perl, Ruby) [1], but compiled languages are not always practical.

2. Monitor usage
   To reduce energy usage of code, a first step would be to monitor energy usage of the system and compare energy usage between different versions of code. For PCs and laptops, directly measuring wattage can be done with a wattmeter placed between the computer and the socket. One drawback is that energy consumption by all processes on the system is measured, which also includes other processes not related to the code under scrutiny. For dedicated servers this method makes more sense and there are also tools developed to measure energy consumption of specific PCI cards, such as General Processing Unit (GPU) boards [2]. Alternatively, libraries exist to perform this task. For Python code the CodeCarbon library gives estimates of energy consumption, based on output from the processors themselves or estimates in case the processors do not monitor energy usage [L2].

3. Use GPUs
   GPUs use more energy than CPUs, but have more computational power and can therefore be more efficient. To maximise efficiency, it is important to properly parallelise the tasks on a GPU. A library such as Kernel Tuner [L3] can take care of this job.

4. Recalculate rather than retrieve from storage
   For simple calculations, recalculation can be more energy efficient than retrieving a previously stored value from local

storage, RAM, or even cache [3]. The complexity of a calculation and the location of storage are important determinants of whether recalculation or retrieving from storage would be most efficient. Since storage location will differ per system and even depend on other tasks running simultaneously on the same machine, it is difficult to distill any general rules. When energy usage is monitored, it can pay off to compare recalculation versus retrieving from storage.

5. Choose a green cluster
Some computing clusters are much more efficient than others. Even within the top ten of greenest HPC clusters there is a twofold difference in energy efficiency [L4]. Some countries or HPC facilities use energy from carbon neutral sources. Check the Green500 [L4] or use, for instance, the CodeCarbon library [L2] to find a green computing cluster.

6. Optimise where it counts
Optimising software for energy efficiency can be very time consuming, given the large number of factors affecting energy usage and its effect on the carbon footprint. Allocating your time to make projects that are computationally intensive less energy demanding and software that has potentially high usage can pay off.

Links:
[L1] https://kwz.me/hq5
[L2] https://codecarbon.io
[L3] https://research-software.nl/software/kernel-tuner
[L4] https://www.top500.org/lists/green500/2022/06/

References:
[1] R Pereira, et al., "Energy Efficiency Across Programming Languages: How Do Energy Time and Memory Relate?", in Proc. of 10th ACM SIGPLAN SLE, 256-267, 2017.
[2] J.W. Romein, B. Veenboer, "PowerSensor 2: a fast power measurement tool," IEEE ISPASS, 111–113, 2018.
[3] M. Pi Puig, L.C. DeGiusti, M. Naiouf, "Are GPUs non-green computing devices?", JCS&T, 18, 153–159, 2018.

Please contact
Reinder Radersma, NWO-I Digital Competence Center, CWI, The Netherlands
+31 20 5924049
Reinder.Radersma@cwi.nl

Making Scientific Research on Dutch National Supercomputer Energy Efficient

by Sagar Dolas (SURF), Ana Verbanescu (Twente University) and Benjamin Czaja (SURF)

Energy is an emerging topic in the scientific computing ecosystem and is becoming a design point for the future of research. Science relies increasingly on digital research computing as a tool for analysis and experimentation. Exponential increase in demand for computing means that classically designed ICT infrastructure will soon become unsustainable in terms of its energy footprint [1]. We need to experiment with energy-efficient methods, tools, and algorithms and hardware technologies. In the Netherlands, we are working towards zero energy waste for high performance computing (HPC) applications on the national supercomputer "Snellius". It involves discussing challenges, proposing new research directions, finding opportunities to engage the user community, and taking steps for responsible use of software in research.

Traditionally, supercomputing focuses on improving latency or throughput, which are of massive importance for applications such as drug discovery or climate simulations. For many decades we developed infrastructure, algorithms, and software tools to obtain improvements. Given the rapid increase in energy usage for ICT services, further emphasised by the imminent energy crisis, it is a priority to understand and optimise the energy consumption of research computing applications [2].

Specifically, our initiative is about working with three stakeholders that need to collaborate to reduce the energy impact: application developers, system integrators, and system operators:
1. *Application developers* are responsible for improving the energy efficiency of their own code, making use of algorithmic, programming, and hardware tools at their disposal. Ideally, applications should be able to adapt to the available system resources and use them effectively. Research into programming models and tools that enable such flexibility is accelerating.
2. *System integrators* are responsible for offering the right resources for the application developers and system operators. These resources must include efficient hardware – e.g., different GPGPUs, CPUs (Central processing unit), or even FPGAs (Field Programmable Gate array) – to enable different application mixes. Research into procuring systems and provisioning applications with the right resources is mandatory.
3. *System operators*, with their holistic view, are responsible for efficiently scheduling workloads on system resources and potential energy harvesting where resources/systems are massively underutilised. Research into tools for energy-efficiency resource management and scheduling, as well as energy harvesting, is ongoing.

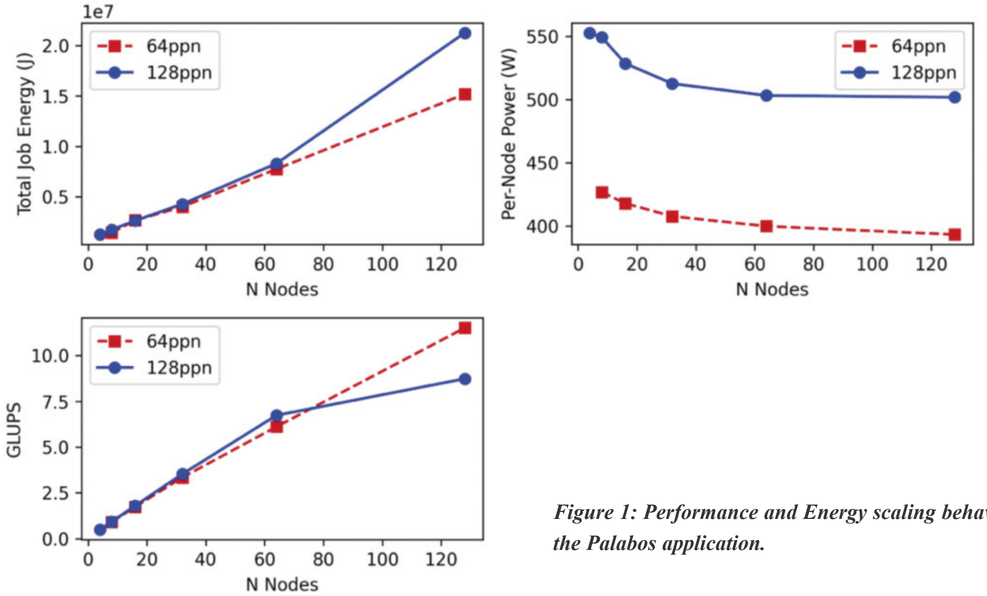


Figure 1: Performance and Energy scaling behaviour of the Palabos application.

To address the interests and concerns of all these stakeholders, we follow three lines of actions:

1. Define methods for application characterisation, towards building a detailed application signature in terms of resource utilisation, performance, and energy consumption.
2. Co-design system-level tools and platforms to allow operators to formulate recommendations to optimise overall energy consumption, and further shape green computing policies for the supercomputing systems.
3. Co-design frameworks to assess and configure systems procurement and resource provisioning for high-efficiency, low-waste application deployment.

Case study

We performed a strong scaling study on the computational fluid dynamics solver Palabos [L1], which is based on the Lattice Boltzmann Method. Palabos, we believe, serves as a typical use case of a memory-intensive HPC application and through performance/energy analysis can serve as a template for energy usage for other similar HPC applications.

As shown in Figure 1, the application scales linearly with increasing the number of nodes. We observe a flattening of the scaling behaviour on larger node counts and identify that the memory bandwidth limits the code. By lowering the processes per node (ppn) of the application, we could maximise the memory bandwidth available to the application, thus resulting in much lower energy usage. The analysis on Palabos represents the value of including energy as a metric and traditional performance analysis. It explains how the resources of a cluster can be adjusted for an HPC application to maximise performance and minimise energy usage.

Also, as shown in Figure 2, we are able to get iteration level information of the application using the system level tool Energy Aware Runtime [3]. With this fine-grained information, we can profile the application in order to identify when and where the application is using the most energy. In Figure 2, we identify that Palabos uses the most energy during the "simulation" phase, where the value of GFLOPS is also very high.

The analysis on Palabos represents the value of including energy as a metric and traditional performance analysis. It explains how the resources of a cluster can

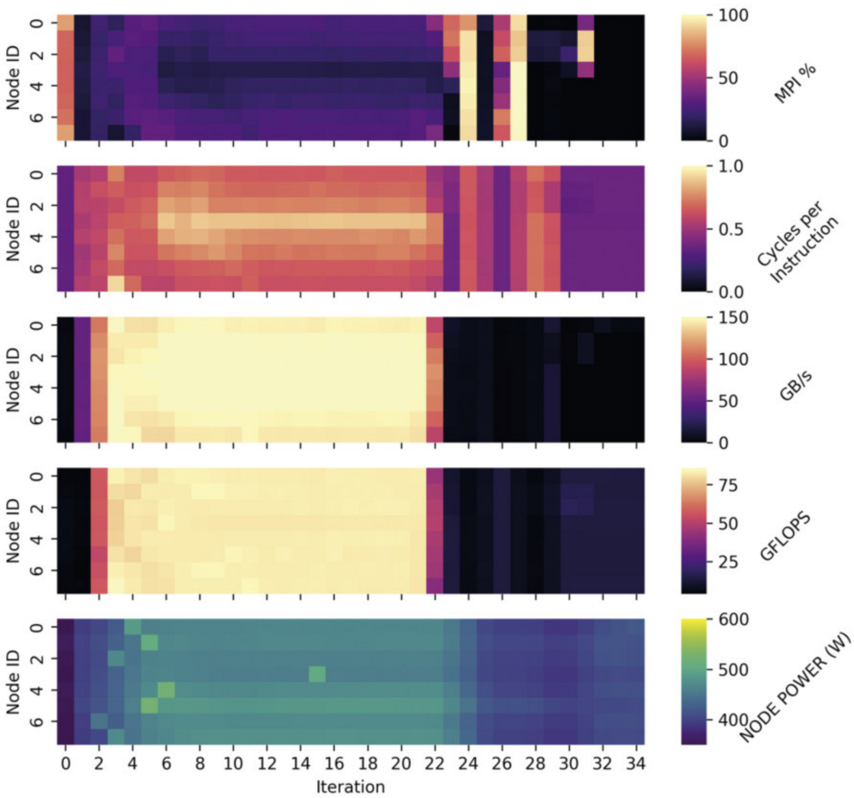


Figure 2: Heatmap analysis of Palabos.

be adjusted for an HPC application to maximize performance and minimize energy usage.

Finally, to address the scientific software community at large, transparency is necessary to assess the energy footprint. To this end, methods, tools, and metrics are mandatory to determine the operation of supercomputers and HPC centres in terms of energy consumption and environmental impact. Our initiative into energy efficiency aims at zero waste and energy awareness for software in research.

Links:
[L1] https://palabos.unige.ch
[L2] https://www.eas4dc.com
[L3] https://kwz.me/hq6
[L4] https://www.surf.nl/en/energy-smart-computing

References:

[1] Dutch Data Centers Association, "State of the dutch data centers - annual report", 2021. https://www.dutchdatacenters.nl/en/publications/annual-report-2021/
[2] European Commission, "A European Green Deal", https://ec.europa.eu/info/strategy/priorities-2019-2024/european-green-deal_en.
[3] J. Corbalan, L. Brochard, "EAR - Energy Management framework for Supercomputers". https://kwz.me/hq7

Please contact:
Sagar Dolas, SURF, The Netherlands
sagar.dolas@surf.nl

Why Might We Fail to Develop Ethical AI/AS?

by Philippe Valoggia (Luxembourg Institute of Science and Technology)

Ethical AI/AS are systems that do not infringe human rights. We might fail to develop such systems due to a lack of common understanding of what quality requirements to meet (i.e., the rights and freedoms of users), and because actors involved in the systems' development do not collaborate in a consistent way. We propose a risk-based engineering approach to overcome these two engineering pitfalls.

Advanced digital technologies such as Artificial Intelligence and Autonomous Systems (AI/AS) open up promising opportunities to tackle economic, societal, and environmental challenges. But their use also gives rise to certain concerns related to security, privacy, and ethics. To prevent privacy issues, the EU General Data Protection Regulation (GDPR) [L1] introduced the obligation to engineer personal data-based products and services that deliberately protect the rights and freedoms of individuals (Data Protection by Design and by Default, Art. 25).

Protection of human rights is a critical quality requirement of any system or product handling personal data. Field observations and literature reviews raise doubt on the ability of current privacy engineering approaches to properly support the development of privacy-preserving products and systems. Literature related to privacy engineering usually adopts a goal-oriented approach to specify privacy quality requirements of a system (see Figure 1).

Protection of rights and freedoms is achieved throughout the completion of privacy-protection goals. Confidentiality, security, integrity, unlikability, transparency, autonomy, etc. have become usual privacy quality requirements of systems handling personal data. The practicability of these goals is sometimes questionable, but more importantly, the measure of their achievement is still not well established, and their respective contribution to the protection of human rights is either hazardous or not specified.

Data protection through technology development is usually presented as a multidisciplinary challenge: it requires combining different fields of knowledge to properly meet privacy quality requirements. By breaking down the protection of human rights into several goals, these approaches implicitly introduce a division of work: experts involved in the design process are likely to specialize in the achievement of only a single or a couple of goals. Experts then operate in silos, and their attention is drawn to the satisfaction of one privacy quality requirement rather than to the protection of people.

The first works related to ethics and AI seem to favour a goal-oriented engineering approach. Indeed, various AI principles are proposed to specify quality requirements of trustworthy AI-based systems. The emergence of disciplinary communities specialising in the fulfilment of one of these principles as

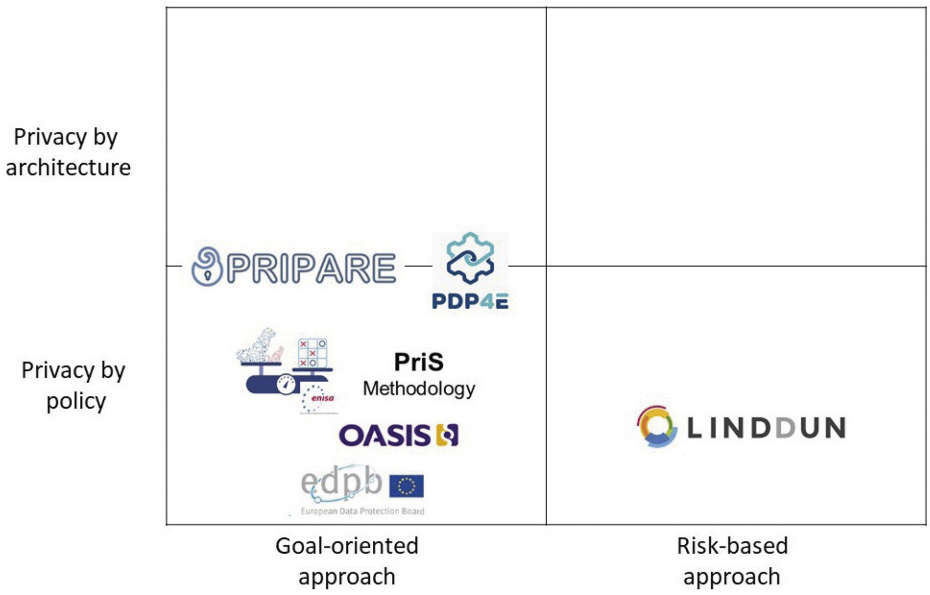


Figure 1: Main privacy-by-design methodologies organised along their approach and their focus.

robustness or explainability has been observed. Ethical engineering is therefore likely to face the same engineering pitfalls as privacy engineering.

Neither privacy engineering nor ethical engineering is doomed to fail to properly protect human rights. It is proposed to adopt a risk-based approach to overcome the two mentioned goal-oriented engineering approach pitfalls. Quality requirement specification consists then of identifying risks that advanced digital systems poses to human rights. From a methodological point of view, a risk-based approach is the key means to specify and to measure critical requirements that cannot be quantified as risks to rights and freedoms of natural persons.

A risk-based approach to specify and measure quality requirements is consistent with the data protection EU legal framework, which is presented as a risk-based regulation. GDPR introduces the obligation to measure the impact of a system on the protection of personal data (Data Protection Impact Assessment – DPIA, art. 35). DPIA is the only recognised indicator of the protection of human rights by the law. It is logical to use it to measure the satisfaction of privacy quality requirements when developing systems handling personal data.

The first proposal of EU's Artificial Intelligence Act seems to adopt a risk-based approach too. Article 9 states the obligation to implement risk management systems for AI-based systems. In addition to its consistency with regulatory framework, a risk-based approach to engineering is likely to minimise the disciplinary silos effect mentioned above. Indeed, considered as a network object (Latour, 2005), risk helps to smooth out disciplinary borders by drawing attention to a common purpose.

Although a risk-based approach to engineering appears as a promising way to tackle the challenges of both privacy and ethical engineering, some investigations are still required to make it happen. It is first necessary to define a risk model that applies to specify risk factors and their interrelations. Second, the assumption that risk is the appropriate network object to ease effective multidisciplinary work when developing advanced technologies-based systems has to be verified. After having designed a privacy risk management assessment tool based on seminal works conducted by Perry [2], we plan to test its reliability and its impacts on multidisciplinary work throughout different privacy and ethical engineering projects, as is suggested by the design science methodological approach.

Links:
[L1] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206
[L2] https://www.list.lu/en/research/

References:
[1] B. Latour, "Reassembling the Social, An introduction to Actor-Network Theory", Oxford University Press, 2005.
[2] S. Perry, "The future of privacy, Private life and public policy", Demos, 1998.

Please contact:
Philippe Valoggia
Luxembourg Institute of Science and Technology, Luxembourg
philippe.valoggia@list.lu

Operational Fairness for Facial Authentication Systems

by Mélanie Gornet (Télécom Paris), Claude Kirchner (CCNE, CNPEN, Inria), and Catherine Tessier (ONERA/DTIS, Université de Toulouse)

How to design a facial authentication system, taking into account both performance and fairness? We consider the choices that a developer makes when coding such a system, such as the training parameters, the architecture of the neural network, or the authentication threshold. We evaluate their impact on the global fairness of the system, showing that fairness is not only affected by the training data but also by the multiple choices that are made when coding the model.

Numerous international recommendations have been issued over the past five years, listing values, principles and criteria to be considered during the development, and more generally the life cycle, of a machine learning system. These recommendations, although paving the way for standardised methods to design algorithms, do not explain how to actually implement these criteria. For example, what should researchers and engineers do to design "fair" machine learning based systems?

We focus on fairness through the eyes of a developer who has to design a facial authentication system. This study was conducted at the French National Committee for Digital Ethics [L1] and is going on as part of a doctoral research. The code is available on GitHub [L2].

Facial Authentication, Performance and Fairness

Automated facial recognition has been particularly criticised for reinforcing overall discrimination that exists in societies. For instance, it was shown that face analysis systems from big tech companies were misclassifying dark-skinned women much more often than light-skinned males [1]. This was later confirmed by the US National Institute of Standards and Technology (NIST) that conducted a study to quantify the accuracy of face recognition algorithms for demographic groups defined by sex, age, and race or country of birth, revealing significant discrepancies [2]. Yet, researchers have long reduced the problem of fairness to a data issue: "Garbage in, garbage out", if the data is unbalanced, the system is quite likely to be biased. But this mindset overlooks other parameters or coding choices that are also likely to affect fairness.

We define digital facial authentication as the comparison of recorded biometric data with those presented by a person. It is a one-to-one matching system, and its output is binary: if the output is "yes", authentication is validated, otherwise it is rejected. We have developed a system using a convolutional neural network (CNN), trained by triplet loss for facial authentication [3]. This process requires many technical choices that are usually made by the developer according to what yields the best performance. We have investigated seven of these choices (see Figure 1) through several metrics for both performance and fairness.

For model selection, a high performance corresponds to a low validation loss at the end of the training phase. For model validation and evaluation, it also corresponds to a high accuracy, a high triplet learned rate (TLR, a metric measuring how well the system has learned), and low error rates.

Fairness is considered here as having the same probability of being recognised by the system in similar conditions, whoever you are. This implies checking, as the NIST did, that for different subgroups of population the system has the same accuracy, TLR and error rates (group fairness). A discrepancy between two groups is significant if the 90% confidence intervals on a given metric do not overlap.

Study Results

Data processing:
- Surprisingly, the data sampling method that yields the best results for fairness measures is the random one, compared to the model prioritising certain underrepresented individuals.

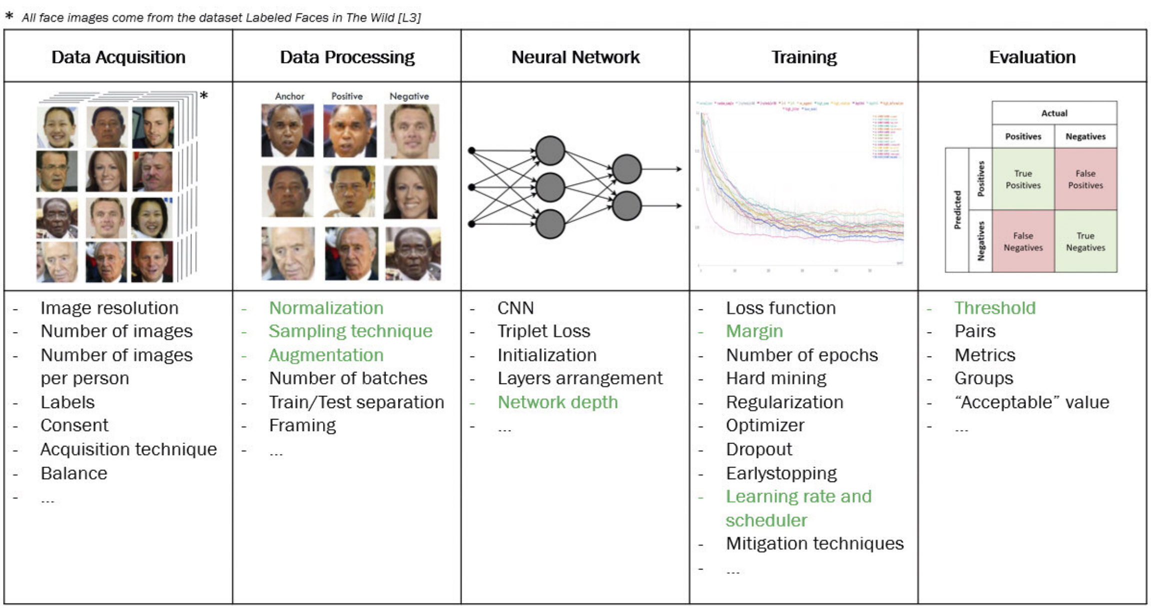


Figure 1: List of design choices for a facial authentication system and investigated choices (in green).

- Data normalisation seems to degrade measures on majority groups but does not affect minority groups.
- Data augmentation improves performance including for minority groups but widens gaps between groups; if data is not augmented, there are fewer gaps but to the detriment of performance.

Neural network:
- The depth of the network does not seem to affect fairness very much but still affects performance.

Training:
- Changing the margin of the loss function can improve fairness but results in a small reduction in performance.
- The choice of the learning rate and its scheduler can affect the local optimum the network will reach and thus yield very different results; here, the model that has the best performance is also the best for fairness.

Evaluation:
- The authentication threshold that separates positive and negative pairs strongly affects the error rates: a high threshold increases the number of matches but generates more false matches, whereas a low threshold prevents some people from being correctly identified. The value of the threshold should thus depend on the use case and on what type of error is the less harmful to the people involved.

Trade-offs

International recommendations about "the ethics of AI" hardly mention that all the proposed criteria cannot be met at the same time and that trade-offs are often necessary. Moreover, fairness is not only a data issue but involves the coding of the model itself. Therefore, ethical thoughts involving all the stakeholders should come with the design of machine learning systems, making the conflicts explicit and guiding the decisions concerning the code implementation as well as the main decision of whether or not to deploy such digital processes.

Links:
[L1] https://www.ccne-ethique.fr/
[L2] https://github.com/mgornet/CNPEN
[L3] http://vis-www.cs.umass.edu/lfw/

References:

[1] J. Buolamwini, T. Gebru, "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification", in Conf. on Fairness, Accountability and Transparency, p. 77–91, PMLR, January 2018.
[2] P. J. Grother, M. L. Ngan, K. K. Hanaoka, "Face Recognition Vendor Test Part 3: Demographic Effects", Technical report, December 2019.
[3] F. Schroff, D. Kalenichenko, J. Philbin, "Face Net: A Unified Embedding for Face Recognition and Clustering", in IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), p. 815–823, June 2015.

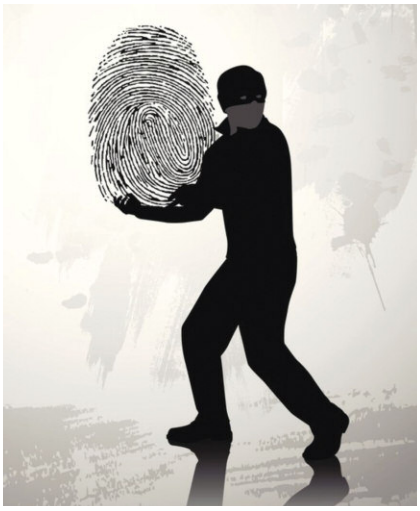
Please contact:

Mélanie Gornet
Télécom Paris, France
melanie.gornet@telecom-paris.fr

Enhancing Biometric Data Security by Design

by Bernhard Strobl (Austrian Institute of Technology, AIT) and Margherita Natali (United Nations)

This article will give an insight into some key problems and related solutions concerning the implementations of a privacy-preserving biometric matching system. We propose three by-design possibilities, strictly in compliance with human rights and data protection regulations, to improve the security of authentication systems: contactless fingerprint scanning, use of a distributed ledger system for biometric matching, and homomorphic encryption. These technical solutions would potentially constitute a step forward for governmental use of authentication procedures under the international security agenda while supporting ethically aligned design principles.



Identity management represents one of the key items on national and international security agendas. In the private domain the use of own identity is predominantly used for granting access in basic and common transactions or actions, whereas governments, in the public domain, more often implement such systems to manage social phenomena such as migration or the illicit activities of organised criminal groups. One of the most common uses of identity management on the global scale is the authentication of official identification documents (e.g., identity cards, passports, driver's licences, and other civil-registry-issued certifications) to monitor and facilitate the legitimate movement of individuals.

Authentication processes can be built upon three basic and very distinct pillars:
- What is known (password, passphrase, PIN, etc.)
- What is available (key, card, stick, document, QR Code, sign, etc.)
- Who the person is (biometrics: DNA, face, fingerprint, iris, veins, etc.)

Sometimes a combination of these pillars is chosen to perform a secure authentication. Depending on the application, different interests may shape the technological choice. For instance, in the case of a commercial service, the need for a speedy and seamless process is prioritised over other interests. Under the public agenda, the authentication procedures used may vary among, for instance, the granting of socio-sanitary services, the exercising of civil rights (e.g., voting), and the implementation of security policies (e.g., preventing terrorism and organised crime, etc.).

The development of advanced biometric technologies that can ensure compliance with Human Rights and data protection regulations and offer reliable outputs, represents an imperative and an opportunity to increase the efficiency of and trust towards the use of authentication procedures.

The scope of this article is to discuss three possible technical solutions to the threats posed to authentication procedures and assess the impact of their implementation by governments in the field of international security [1].

Two of the most threatening attack vectors addressing governmental biometric authentication systems are:

1. Presentation Attack (PA): The attacker presents stolen or replicated biometric samples to the acquisition system to perform the authentication. In order to contrast these attacks some PA-detection (PAD) techniques are available. Another form of face spoofing is embodied by morphed pictures (morphing), which are purposefully difficult to distinguish from the original for the officer or the system performing the authentication.
2. Infiltration and data interception of server systems: A biometric database is leaked, either by directly hacking the server system or by intercepting the data in the transmission path. In both cases, a potential attacker gets access to biometric data that can be duplicated, counterfeited, illicitly stored, analysed, etc.

A valid model to overcome such threats:

a. is a combined approach of reliable systems
b. can perform the authentication through a distributed ledger
c. can be used in combination with homomorphic encryption methods.

As the current tests show, contactless biometric systems (a), those using fingerprints in particular, have over a 98% rate of successful performance [2]. The efficiency of these systems relies on their neutrality of performance towards dry or wet fingers resulting in the collection of high-quality images; and the fast response, which ultimately improves the end-user experience. Compared to previous touch-based technologies, a three-dimensional spoof of accurate fingerprint minutiae data for four fingers presented at once, is extremely hard to achieve. Additionally, PAD methods for fingerprint spoofs in the 3D domain are much easier to detect. Under this perspective, the possibility of a more accurate collection and sharing of biometrics performed by designated authorities under the national security agenda and in full respect for human rights and applicable data protection regulations, could offer wide potential for improvement of the interoperability and systematisations between national systems and international dedicated databases.

The correlation between technologies and Human Rights is more often at the centre of the debate surrounding the use of biometrics. A biometric matching service, where a server system – by design – cannot be compromised thanks to a distributed ledger system, constitutes a tremendous step forward to a more secure and reliable process (b). Trust, according to these features, is built by using several computational nodes/ledgers verifying a “transaction” located at different premises. Compromising one system would trigger alarms. Such architecture, especially if implemented in a multi-party computational matching system, is by-design offering greater guarantees that the biometric data is processed in an exhaustive privacy-preserving manner. Accordingly, the concrete data-processing actions are performed on a fragmented part of the data and none of the servers nor the transmission lines reveal or have access to the entirety of the data. In this sense, such a feature, if adopted by national entities in their biometric authentication procedures and exchange of outputs with the international community, would exponentially increase the compliance of such mechanisms with human rights and data protection regulations, which, in turn, will also increase their reliability and related social confidence.

Finally, studies show that the most secure privacy-by-design principle would be the use of homomorphic encryption (c) [3], which would decrease to the minimum the risks of infiltration of the server and, in combination with the other two above-described features, would eradicate the risk of errors in authentication procedures. Therefore, it could be concluded that technologies, together with adequate policy and legal frameworks, could not only support the work of governments to maintain security and promote Human Rights, but also facilitate it through respective design.

References:

[1] United Nations Security Council Resolutions 2396 (2017) and 2482 (2019), in particular, with Res. 2396 (2017) Member States are required by the UN Security Council to adopt national biometric systems to monitor the cross-border movement of people.
[2] A. Torres, “Interoperability Assessment 2019: Contactless-to-Contact Fingerprint Capture”, NISTIR 8307. https://nvlpubs.nist.gov/nistpubs/ir/2020/NIST.IR.8307.pdf
[3] W.A.A. Torres, N. Bhattacharjee, B. Srinivasan, “Privacy-preserving biometrics authentication systems using fully homomorphic encryption”, in Int. Journal of Pervasive Computing and Communications, Vol. 11 No. 2, pp. 151-168, 2015. https://doi.org/10.1108/IJPCC-02-2015-0012

Disclaimer:

The contribution to this article was made – free of compensation - in Margherita Natali’s private capacities as a researcher and subject matter expert on the legal frameworks applicable to the processing of biometric data for public purposes. The reference materials used in this publication, including their respective citations, do not imply the expression of any opinion whatsoever nor the official endorsement by the United Nations. Any information that may be contained in this publication emanating from actions and decisions taken by States does not imply recognition by the United Nations of the validity of the actions and decisions in question and is included without prejudice to the position of any Member State of the United Nations.

Please contact:

Bernhard STROBL, Austrian Institute of Technology (AIT)
Bernhard.strobl@ait.ac.at

Rethinking Requirements Engineering for Sustainability

by Alessio Ferrari (CNR-ISTI) and Manlio Bacco (CNR-ISTI)

We present a new paradigm of requirements engineering research for sustainability. This study proposes to go beyond stakeholders' goals, and introduces the concepts of drivers, barriers and impacts of technology in a certain domain. We collect information about these constructs in an interview study with 30 experts on digitalisation in forestry, agriculture, and rural areas.

Sustainability in system engineering has traditionally been interpreted as the ability of a system to evolve and be maintained in a cost-effective way, while managing technical debt. This vision, which focuses only on the technical side of sustainability, has been criticised by the Karlskrona Manifesto [1], edited by a group of software engineering researchers to raise awareness on the relationship of Information and Communications Technology (ICT) solutions with ecological and social systems. The manifesto calls for a more systemic view of sustainability during system design, and identifies requirements engineering (RE) as the key area where system-level thinking can be applied to escape the trap of solutionism and broaden the perspective to reason on potential effects of technological change from the social, ecologic, ethical, and economic viewpoints.

The call to arms of the Karlskrona Manifesto triggered research around the notion of sustainability requirements. These are intended as quality goals that a system will fulfill to provide long-term benefits for its environment and members therein, while minimising damage to other members and the environment as a whole. Different RE approaches have been proposed to elicit this particular type of requirement. Part of them focus on energy-management aspects, and use different combinations of RE practices – prototyping, design thinking, goal modelling, etc. – specifically tailored to elicit requirements concerning the energy-efficiency of the system. Others take a domain-agnostic perspective and propose general sets of sustainability requirements patterns, interview scripts, as well as guidelines to rethink the software process considering sustainability as a main concern. However, existing strategies lack the proper high-level view to deal with the societal and long-term impacts of the transformation entailed by the introduction of a new technological solution.

Our research, conducted within the framework of the H2020 DESIRA (Digitisation: Economic and Social Impacts in Rural Area) project [L1], proposes to go beyond the concept of system requirements and stakeholders' goals and raise the degree of abstraction by focusing on the notions of drivers, barriers, and impacts that a system can have on the environment in which it is deployed [2]. Drivers include goals of some stakeholders, for example, the need to improve the quality of products or processes by certain actors, but also other higher-level aspects, for example, the funding from institutions to support specific technologies. Barriers are intended as elements preventing the achievement of a specific goal, but are also more structural impediments that hamper the introduction of the digital technology as a whole in the given context. For example, the difficulty of certain actors in interacting with the novel technology, or the regulatory problems related to the use of a certain technology that does not account for privacy issues.

The concept of impact is intended as the expected effect that the digital technology can have from a sustainability standpoint, and thus in mid- to long-term. The impact can be positive, as, e.g., reduction of manual labour, but also negative, for example, due to the exclusion of subjects who cannot afford the technology. The informal meta-model depicting the relationships between these concepts is reported in Figure 1.

To put this vision into practice, we interview 30 cross-disciplinary experts in the representative domain of rural areas, and

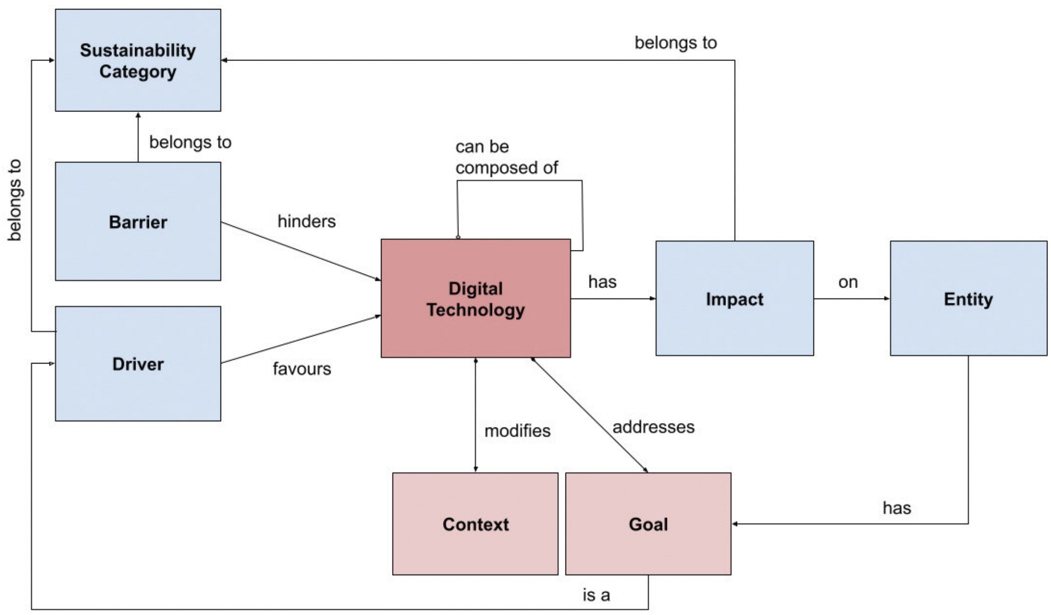


Figure 1: Meta-model representing the relationships between the concepts introduced in the study.

we analyse the transcripts to identify common themes. As a result, we provide drivers, barriers, and positive or negative impacts associated with the introduction of novel technical solutions in rural areas. Our results show that typical barriers to the adoption of ICT solutions are the lack of connectivity in rural areas, but also fear and distrust towards technology. In addition, the cost of technology and regulatory issues, also related to unclear data governance are relevant barriers.

The main drivers are economic, as technology can lead to cost reduction, but also ecological and institutional, since technology can improve monitoring as well as accountability. In this regard, regulators can play a crucial role by means of funding programs and norms. Positive impacts are the replacement of repetitive labour and the possibility of exploiting economies of scale. On the other hand, negative impacts are the higher dependency on technology as well as the social exclusion of some players that cannot cope with the change, at least not fast enough. This work contributes with a paradigm shift in the analysis of sustainability requirements, by introducing the concepts of drivers, barriers, and impacts associated with the adoption of technological solutions. Furthermore, our themes represent a reliable snapshot of the situation in rural areas and can be taken as a reference for the development of socio-technical systems in this domain.

Link:
[L1] https://desira.d4science.org/

References:
[1] C. Becker, et al., "Sustainability design and software: The karlskrona manifesto", in 2015 IEEE/ACM 37th IEEE International Conference on Software Engineering (Vol. 2, pp. 467-476), IEEE 2015. DOI: 10.1109/ICSE.2015.179
[2] A. Ferrari, et al., "Drivers, barriers and impacts of digitalisation in rural areas from the viewpoint of experts", Information and Software Technology, 145, 106816, 2022. DOI: 10.1016/j.infsof.2021.106816

Please contact:
Alessio Ferrari, CNR-ISTI, alessio.ferrari@isti.cnr.it
Manlio Bacco, CNR-ISTI, manlio.bacco@isti.cnr.it

Ethics and Human Aspects in Pandemic Management

by Karin Rainer (AGES), Viktoria Fischer (AGES), Alois Leidwein (AGES), and Georg Neubauer (AIT)

Digital solutions and support systems in pandemic management are the next step in enhancing current structures and overcoming identified gaps in the current state of the art. Currently, decision support systems are often rudimentary and seldom include the ethical and user-centric needs of the practitioners they are intended for. Thus, the "ROADS to Health" approach can give new insights focusing on requirements and pro-actively addressing and including ethical as well as holistic human aspects for enhanced applicability and user acceptance.

Looking back, most European countries were inadequately prepared for a pandemic of the dimension and manifestation of COVID-19. The uncertain epidemiological – but also infrastructural – information landscape made decision-making regarding public health strategy and targeted mitigation measures difficult, fluid, and highly disputed [1]. This also made it clear that current digital support tools were barely apt to answer the multiple, complex, and dynamic needs and requirements of the decision makers and of health authorities.

The often hard-learned lessons of the last years led to the conclusion that tailor-made systems and modules are required to generate supporting information and insight for an enhanced, evidence-based, and transparent decision support. Such a tool must also consider the requirements of the end users. In addition, the perspective and compliance of the "subjects" in regards to the measures and interventions (e.g., lockdowns) were not sufficiently considered, and it became evident that they need to be included in the development of future decision-support tools. Also, structured identification of specific requirements and pandemic parameters was clearly an important step in the enhancement of pandemic management (see Figure 1).

Thus, the project "ROADS to Health", a project funded under the national KIRAS security research program, is focusing on generating the basis for the matching of mitigation measures and requirements for tackling pandemics/epidemics from different angles, and including a broad range of specific needs. In this regard, epidemiologic, logistic, and strategic aspects, as well as the definition of other pandemic parameters like interventions and resources like equipment, can be identified as a relevant basis for a targeted intervention selection and measure matching following the current evidence.

Ethical aspects and holistic approaches to human needs like usability and optimised user interface design are only two aspects of the broad range of requirements that have to be integrated in an acceptable and actively used support system. After an initial concept to enhance the pandemic management (Rainer et al. 2021), it also became evident that international, scientific "lessons learned" from the COVID-19 pandemic as well as the active integration of stakeholders and end users and

Figure 1: Structuring of pandemic parameters.

of ethical, gender-related, and legal issues have to be the strong backbone for any such endeavour.

With this background, the alignment of physical, interaction, and procedural measures with specific requirements answers the dynamic needs and manifestations of the pandemic and can lead to acceptable and transparent interventions that are also carried by the community. The ethical acceptability and inclusion of human factors is key for the further application of any support tool in a social system. This encompasses, for example, the inclusion of marginalised groups, aspects of the (e.g., age-related) digital divide, gender-related issues, and specifics but also the side-effects of impeded measures. These could be the long-term psychological effects and socio-economic burdens that measures have on the population or on parts of it.

The current effects of the lockdowns in schools on the psychological well-being and development of children and adolescents are another of the many examples of systemic effects of pandemic mitigation measures that have to be taken into account in any support tool for informed decision-making support.

These “collateral damages” and negative side-effects from social deprivation, delay in medical treatment, preventive measures, and many more aspects due to some COVID-related interventions (e.g., closing or reduction of resources in hospitals and general practitioners) must be considered in relation to the intended outcomes of the mitigation of the pandemic [2]. This difficult and complex balancing in a heated discussion at the beginning but also in the course of a pandemic event is a key communication and discourse challenge. Risk communication and the building of a stable ethical framework and trust in the official stakeholders is key to achieve a sustainable acceptance of the impeded interventions.

Finally, the ethical dimension also impacts on a discourse layer the strategic dimension of a holistic measure matching tool. Any strategy, be it “Zero COVID”, “Flattening the Wave”, or keeping the medical system from overload, has to be linked with ethical and human factor implications. The question of saving lives, of secondary effects, of seclusion and infringement of rights has to be a core aspect of any technological solution in this complex area. Developers and researchers must be very aware of these factors to foster applicable, acceptable, and helpful support tools and applications for decision-makers to support the stability and safety of a community.

References:

[1] E. Abrams, S. Szefler, “COVID-19 and the impact of social determinants of health”, in The Lancet Respiratory Medicine, Vol. 8, Issue 7, 647-742, 2020.
[2] D. Lewis, “What Scientists have Learnt from COVID Lockdowns”, in Nature, Vol 609, 236-239, 2022.
[3] K. Rainer, et al., “Risk, Vulnerability and Resilience Analyses for Pandemic Management: Challenges and Perspectives of Approaches for the Future”, IDIMT-2021, Pandemics: Impacts, Strategies and Responses, 29th Interdisciplinary Information Management Talks, 163-174, ISBN 978-3-99113-261-5, Trauner Verlag. 2021.

Please contact:

Karin Rainer
AGES - Austrian Agency for Health and Food Safety, Austria
karin.rainer@ages.at

Plausible Reasoning that Mimics Human Argumentation

by Dave Raggett (W3C/ERCIM)

Introduction

Machine reasoning has had little attention over the last decade when compared to knowledge graphs and deep learning. Application logic is usually buried in the program code, making it cumbersome and costly to update. Approaches based upon traditional logic fail to cope with everyday knowledge that inevitably includes uncertainty, incompleteness and inconsistencies, whilst statistical approaches are often impractical given difficulties in obtaining the required statistics.

Plausible reasoning, by contrast, seeks to mimic human argumentation in terms of developing arguments for and against a given premise, using a combination of symbolic statements and qualitative metadata. Let's start by considering what is meant by knowledge, and its relationship to information and data. Data is essentially a collection of values, such as numbers, text strings and truth values. Information is structured labelled data, such as column names for tabular data. Knowledge is understanding how to reason with information. Knowledge presumes reasoning and without it is just information. As such, it makes sense to focus on automated reasoning for human-machine cooperative work that boosts productivity and compensates for skill shortages.

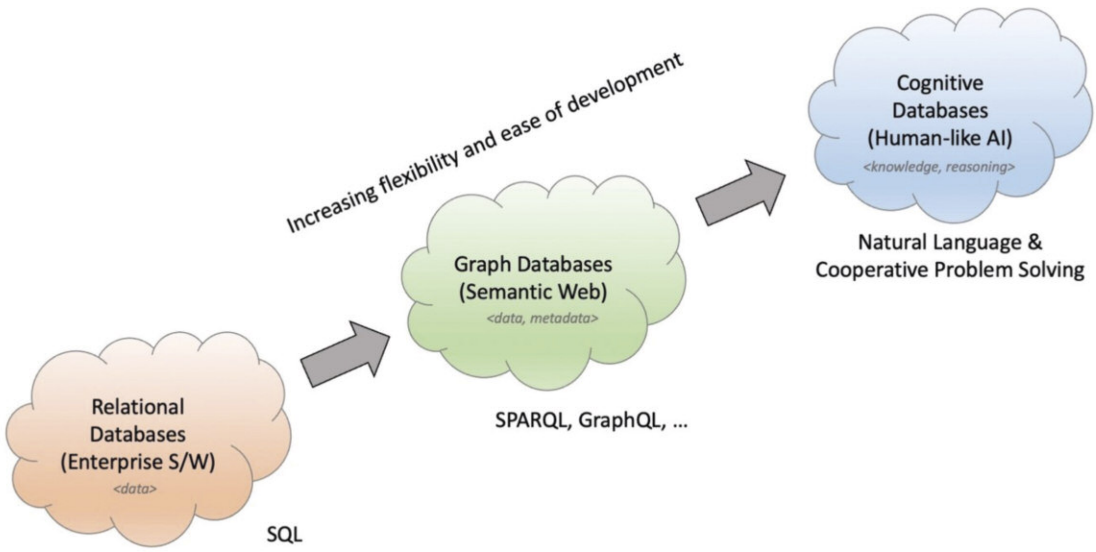
Business software is for the most part based on relational databases that represent data in terms of tables. There is growing interest in the greater flexibility of graph databases using RDF or Property Graphs. The next stage is likely to see the emergence of cognitive databases featuring human-like reasoning along with support for natural language interaction and multimedia rendering (Figure 1).

Plausible Reasoning

People have studied the principles of plausible arguments since the days of Ancient Greece, e.g., Carneades and his guidelines for effective argumentation. There has been a long line of philosophers working on this since then, including Locke, Bentham, Wigmore, Keynes, Wittgenstein, Pollock and many others.

Plausible reasoning is everyday reasoning, and the basis for legal, ethical and business decisions. Researchers in the 20th century were side-tracked by the seductive purity of mathematical logic, and more recently, by the amazing magic of deep learning. It is now time to exploit human-like plausible reasoning with imprecise and imperfect knowledge for human-machine cooperative work using distributed knowledge graphs. This will enable computers to analyse, explain, justify, expand upon and argue in human-like ways.

In the real world, knowledge is distributed and imperfect. We are learning all the time, and revising our beliefs and understanding as we interact with others. Imperfect is used here in the sense of uncertain, incomplete and inconsistent.



Conventional logic fails to cope with this challenge, and the same is true for statistical approaches, e.g., Bayesian inference, due to difficulties with gathering the required statistics. Evolution has equipped humans with the means to deal with this, though not everyone is rational, and some people lack sound judgement. Moreover, all of us are subject to various kinds of cognitive biases, as highlighted by Daniel Kahneman.

Consider the logical implication $A \Rightarrow B$. This means if A is true then B is true. If A is false then B may be true or false. If B is true, we still can't be sure that A is true, but if B is false then A must be false. Now consider a more concrete example: if it is raining then it is cloudy. This can be used in both directions: Rain is more likely if it is cloudy, likewise, if it is not raining, then it might be sunny, so it is less likely that it is cloudy, which makes use of our rough knowledge of weather patterns.

In essence, plausible reasoning draws upon prior knowledge as well as on the role of analogies, and consideration of examples, including precedents. Mathematical proof is replaced by reasonable arguments, both for and against a premise, along with how these arguments are to be assessed. In court cases, arguments are laid out by the Prosecution and the Defence, the Judge decides which evidence is admissible, and the guilt is assessed by the Jury.

During the 1980's Alan Collins and co-workers developed a core theory of plausible reasoning based upon recordings of how people reasoned aloud [R1]. They discovered that:
1. There are several categories of inference rules that people commonly use to answer questions.
2. People weigh the evidence bearing on a question, both for and against, rather like in court proceedings.
3. People are more or less certain depending on the certainty of the premises, the certainty of the inferences, and whether different inferences lead to the same or opposite conclusions.
4. Facing a question for which there is an absence of directly applicable knowledge, people search for other knowledge that could help given potential inferences.

A convenient way to express such knowledge is the Plausible Knowledge Notation (PKN). This is at a higher level than RDF, and combines symbols with sub-symbolic qualitative metadata. PKN statements include properties, relationships, dependencies and implications. Statements may provide qualitative parameters as a comma separated list in round brackets at the end of the statement.

Qualitative metadata is used to compute the degree of certainty for each inference, starting from the certainty of the known facts, and using algorithms to combine multiple sources of evidence:
- typicality in respect to other group members, e.g., robins are typical song birds;
- similarity to peers, e.g., having a similar climate;
- strength, inverse – conditional likelihood in each direction, e.g., strength of climate for determining which kinds of plants grow well;
- frequency – proportion of children with given property, e.g., most species of birds can fly;
- dominance – relative importance in a given group, e.g., size of a country's economy;
- multiplicity – number of items in a given range, e.g., how many different kinds of flowers grow in England.

How does this support reasoning? Let's start with something we want to find evidence for
flowers of England includes daffodils
and evidence against it using its inverse:
flowers of England excludes daffodils
We first check if this is a known fact and if not look for other ways to gather evidence.
We can generalise the property value:
flowers of England includes ?flower
We find a matching property statement:
flowers of England includes temperate-flowers
We then look for ways to relate daffodils to temperate flowers:
daffodils kind-of temperate-flowers
Allowing us to infer that daffodils grow in England.
Alternatively, we can generalise the property argument:
flowers of ?place includes daffodils
We look for ways to relate England to a similar country:
Netherlands similar-to England for flowers
We then find a related property statement:
flowers of Netherlands includes daffodils, tulips.

This also allows us to infer that daffodils grow in England. The certainty depends on the parameters, in this case "similarity". These examples use properties and relationships, but we can also look for implications and dependencies, e.g., a medium latitude implies a temperate climate, which in turn implies temperate flowers. We can prioritise inferences that seem more certain, and ignore those that are too weak.

A proof-of-concept web-based demo can be found in [L1]. It introduces the plausible knowledge notation and applies it to a suite of example queries against a cognitive knowledge base, including reasoning by analogy and the use of fuzzy quantifiers, something that is needed to support the flexibility of natural language, e.g., none, few, some, many, most and all, as in:
Are *all* English roses red or white?
all ?x where colour of ?x includes red, white from ?x kind-of rose and flowers of England includes ?x
Are only a *few* roses yellow?
few ?x where colour of ?x includes yellow from ?x kind-of rose
Which English roses are yellow?
which ?x where colour of ?x includes yellow from ?x kind-of rose and flowers of England includes ?x
Are *most* people older than 20?
most ?x where age of ?x greater-than 20 from ?x isa person
Is *anyone* here younger than 15?
any ?x where age of ?x less-than 15 from ?x isa person
How many people are slightly younger than 15?
count ?x where ?x isa person and age of ?x slightly:younger-than 15
How many people are very old?
count ?x where ?x isa person and age of ?x includes very:old.

Plausible reasoning embraces Zadeh's fuzzy logic in which scalar ranges are described as blend of overlapping values for imprecise concepts like warm and cool. This enables simple control rules to be expressed using terms from the ranges. Fuzzy sets correspond to multiple lines of argument, e.g., the certainty that the fan speed is stopped, slow or fast. Fuzzy modifiers model adverbs and adjectives, such as very, slightly and smaller, by transforming how terms relate to scalar ranges.

Analogical queries can be solved by looking for structural similarities, e.g.,

leaf:tree::petal:?	short:light::heavy:?
leaf **part-of** tree petal **part-of** flower	short **less-than** long **for** size light **less-than** heavy **for** weight

When comes to scaling to very large knowledge graphs, an attractive approach is to decompose large graphs into overlapping smaller graphs that model individual contexts. Such contexts are needed to support reasoning about past, present and imagined situations, e.g., when reasoning about the future or counterfactual reasoning about causal explanations. Contexts are also needed for natural language, e.g., consider "John opened the bottle and poured the wine". This is likely to be a social occasion with wine being transferred to the guests' glasses, with a context associated with causal knowledge, e.g., to pour liquid from a closed bottle, it first needs to be opened.

Humans find coherent explanations very quickly when listening to someone speaking. One potential mechanism to mimic this is to exploit spreading activation. This can be used to identify shared contexts and the most plausible word senses, as well as to mimic characteristics of human memory such as the forgetting curve and spacing effect. Spreading activation can also be applied to guide search for potential inferences as part of the reasoning process, as noted by Collins.

The web-based demo [L1] uses a small set of static reasoning strategies. Further work is needed to introduce metacognition for greater flexibility, and to reflect the distinction between System 1 & 2 thinking as popularised by Daniel Kahneman in his work on cognitive biases. Work is now underway to demonstrate how short natural language narratives can be understood using plausible reasoning over common sense knowledge, along with the role of metaphors and similes.

This can be contrasted with large language models derived using deep learning in that the latter rely on statistical regularities using opaque representations of knowledge that aren't open to inspection. The ability to explain and justify premises is a clear benefit of plausible reasoning. An open question is how to integrate plausible reasoning with approaches based upon deep learning, e.g. for applying everyday knowledge to improve overall semantic consistency for images generated from text prompts.

A further challenge will be to support continuous learning, e.g., syntagmatic learning about co-occurrence regularities, paradigmatic learning about abstractions, and skill compilation for speeding common reasoning tasks. Learning can be accelerated through the use of cognitive databases shared across many cognitive agents as a form of artificial hive mind, and combined with knowledge derived from large corpora.

This work has been supported through funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 957406 (TERMINET) and No. 833955 (SDN-microSENSE).

Link:
[L1] A web-based proof of concept:
https://www.w3.org/Data/demos/chunks/reasoning/

Reference:
[1] The Logic of Plausible Reasoning: A Core Theory: Allan Collins & Ryszard Michalski, Cognitive Science 13, (1):1-49 (1989).

Please contact:
Dave Raggett, W3C/ERCIM, dsr@w3.org

Dynamic Resource Reservation for Commoditised Space Resources

by Fabrice Saffre (VTT Technical Research Centre of Finland)

Low Earth Orbit (LEO) satellite constellations are fast becoming a reality, but like for many new technologies, making the most out of them may require some innovative thinking beyond the obvious.

There is an ongoing paradigm shift in how orbital assets are being designed and deployed. Whereas it used to be the case that a given mission or specific application (combined with budget constraints) dictated which satellites should be placed onto which orbit, we are increasingly in a situation where potentially "multi-purpose" constellations are being launched with the goal of maximising planetary coverage [1]. Although many constraints and challenges remain, such as the increase in energy consumption and additional onboard equipment required to perform multiple functions, a more versatile space infrastructure would have tremendous advantages. For now, early examples of purpose-built LEO constellations include SpaceX's Starlink [L1] and the European Commission's plan for satellite broadband [L2].

In the hypothetical scenario in which software-defined functions make the "satellite-as-as-service" paradigm feasible, existing resources are already circling the Earth at the time when a new, possibly relatively short-term mission (e.g., monitor the course of a suspicious vessel) is being considered. The challenge is then to select and temporarily allocate the best possible subset of available orbital platforms to complete it. Factor in the possibility of multiple concurrent tasks and conflicting priorities and it may become a complex optimisation problem that requires solving in near real-time, which we're only beginning to comprehend [2].

Extrapolating from previous work [3], we have begun developing a method based on nature-inspired heuristics (including but not limited to genetic algorithms) to identify a target number of satellites among the members of a constellation of arbitrary size/density that would be best suited to collectively perform a particular task. In the selected proof-of-concept, a task is defined as monitoring a given location (latitude and longitude) over a given period (start and end time). In preliminary testing, there are 3 such tasks: 36° N, 20° E (00:00-06:00), 12° S, 12° W (03:00-09:00) and 36° N, 18° W (06:00-12:00).

The arbitrarily set parameters are as follows: the planner has a "budget" of 32 satellites, to be selected in a patchy constellation 256 strong (i.e., one out of eight). The objective is simply to maximise the observation time of all three locations, during their respective (and partially overlapping) time windows. Gaps in coverage (interval between observations) and uneven distribution between tasks are not factored in.

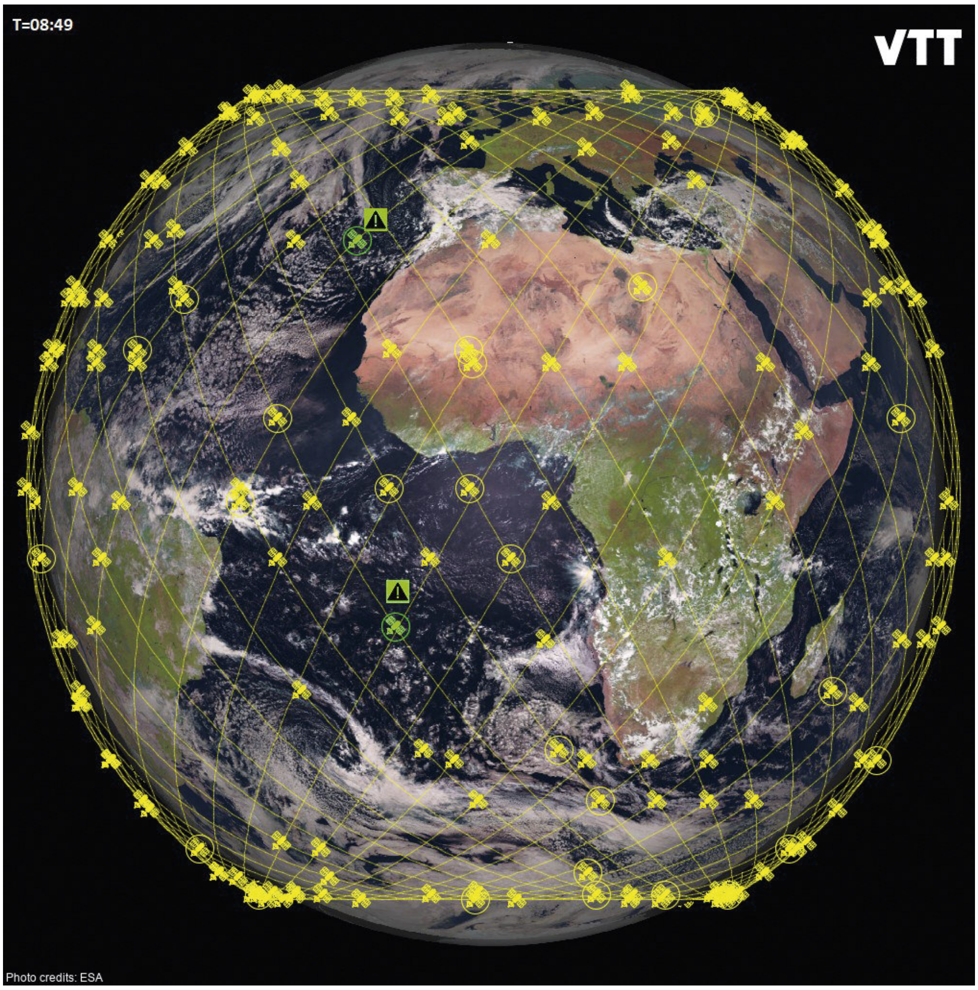


Figure 1: Snapshot of the numerical experiment after heuristic search. Satellites that are circled were selected. A green icon indicates that the satellite was (a) selected, and (b) is above the horizon of one of the designated target locations during the target observation time window ("attention" sign).

After a heuristic search of approximately ten seconds (HP Elitebook, Intel Core i5 8th Gen), observation time averaged ~7h 45' (out of a max. possible 18h). For comparison, a random set of satellites (for the same parameter values) averaged ~2h30' cumulative observation time, i.e., well over 200% improvement (NB: there is a law of diminishing return so a much longer search does not yield substantially better results). Fig. 1 is a snapshot of the simulation identifying the sub-set of satellites that were allocated to the task.

A valid discussion point is whether such heuristics will have practical applications. Orbital mechanics are essentially deterministic, which implies that exact calculation methods may be able to identify a better optimum at a lower computational cost. However, the same cannot be said of the tasks (which may be created dynamically and follow an arbitrary and so unpredictable space-time pattern). Furthermore, it can be anticipated that there will be a complexity barrier above which the number of concurrent tasks and/or other factors affecting the "patchiness" of the constellation (e.g., some satellites may be otherwise engaged, undergoing maintenance etc.) will make the flexibility, simplicity and robustness of our method an attractive proposition.

This is very preliminary work, but it perhaps has the merit to illustrate how unconventional methods from a different field can sometimes yield simple, computationally frugal solutions to a complex problem. An exclusive focus on technical feasibility often delivers a new infrastructure that is functional but remains underutilised, due to some operational aspects having been insufficiently researched. Maybe the proposed approach will reveal itself to be impractical, but such "what-if" questions are the conceptual fuel that innovation runs on.

Links:
[L1] https://kwz.me/hqx
[L2] https://kwz.me/hqz

References:

[1] A. Kak and I. F. Akyildiz, "Designing Large-Scale Constellations for the Internet of Space Things With CubeSats," in IEEE Internet of Things Journal, vol. 8, no. 3, pp. 1749-1768, 1 Feb.1, 2021. https://doi.org/10.1109/JIOT.2020.3016889
[2] T. Leng, et al., "Collaborative Computing and Resource Allocation for LEO Satellite-Assisted Internet of Things," in Wireless Communications and Mobile Computing, vol. 2021, Article ID 4212548, 12 pages, 2021. https://doi.org/10.1155/2021/4212548
[3] H. Hildmann, et al., "Termite algorithms to control collaborative swarms of satellites. International Symposium on Artificial Intelligence, Robotics and Automation in Space" (i-SAIRAS 2018), Madrid, Spain, June 4-7, 2018. Published in: European Space Agency Proc. of i-SAIRAS 2018. Retrieved March 8, 2022. https://kwz.me/hqA

Please contact:

Fabrice Saffre, VTT Technical Research Centre of Finland
fabrice.saffre@vtt.fi

Empirical Software Engineering and Formal Methods for IoT Systems

by Davide Basile, Maurice ter Beek, Giovanna Broccia and Alessio Ferrari (ISTI-CNR)

Researchers from the Formal Methods and Tools (FMT) lab of ISTI-CNR are working on the application of formal methods to devise interaction protocols for safe-by-construction IoT Systems of Systems. They are also working on the empirical investigation and evaluation of the effectiveness of techniques and methodologies proposed for IoT application scenarios. The research is being conducted in the context of the national project T-LADIES, funded by the Italian Ministry of Education, University and Research (MIUR) under the program for Projects of National Interest (PRIN).

T-LADIES (Typeful Language Adaptation for Dynamic, Interacting and Evolving Systems) [L1] targets the development of advanced technologies, formal methods, and tools for reliable and efficient distributed applications in cloud-to-edge-to-IoT systems. Conventional language and engineering techniques struggle to keep up with the pace at which such applications evolve, and they provide inadequate support for development and maintenance, automatic property verification and enforcement, and bug detection.

T-LADIES proposes to use dynamic language adaptation to improve the development and maintenance process in combination with advanced type systems to verify and enforce the properties of software systems. The aim is to increase productivity and software quality, while reducing development costs and time-to-market.

T-LADIES focuses on the Internet of Things (IoT), which is characterised by several interconnected "things", typically in Systems of Systems, which have heterogeneous capabilities and behaviour, and whose spatial distribution is a relevant parameter. These characteristics pose several key challenges:
- IoT applications are programmed in general-purpose, domain-agnostic languages, which hinders their maintenance, modification, and evolution. Since IoT is pervasive, it is desirable that domain experts – without advanced programming skills – can easily (re)program, maintain, and modify IoT applications.

- “Things” are assumed to interact according to predefined schemes, whereas context and application changes call for supporting configurations that are unknown when "things" are conceived. Such flexibility requires richer ways to specify "things" capable of improving the approaches typically offered by general-purpose languages.
- IoT applications are based on the interaction of "things" that dynamically vary in number and kind, which poses issues of correctness, dynamic evolution, and adaptation. The notion of interaction should thus be enriched to support variability, dynamic monitoring, property enforcing, and orchestration of the "things" in the general context of Systems of Systems.
- IoT applications and systems have to meet high-quality standards, like absence of undesired situations (e.g., deadlocks, orphan messages, etc.), support for non-functional requirements (e.g., performance, energy sustainability, etc.), and resilience to varying or reconfigurable execution contexts. The complexity of IoT applications and Systems of Systems makes it hard to meet and preserve such standards without automated tools.

T-LADIES addresses these challenges through an approach to software development that mixes language adaptation, interaction mechanisms, and advanced type systems. In T-LADIES, language adaptation will make it possible to vary how the language behaves in different contexts and, consequently, to modify application behaviour accordingly, with no impact on the source code. Interaction protocols are intended to provide extra functionality to the mechanisms that are natively available in the context in which the application runs. Advanced type systems enable the behavioural specification of entities, the enforcing and verification of system properties, and the early detection of bugs.

The overall goal is to achieve results of both foundational and practical impact. The expected outcome is a novel formal approach to develop and maintain modern applications by focusing on dynamic adaptation, property enforcing, and component interaction. In particular, based on experience in studying compositionality in terms of configurability and adaptability at runtime, and on formal methods and tools for guaranteeing safe communication in Systems of Systems modeled as (contract or team) automata [R1], FMT will introduce variability in multi-party synchronisation type specifications and interaction protocols, as well as criteria to preserve communication properties (e.g., receptiveness, responsiveness, agreement, etc.) by composition, akin to the correctness-by-construction paradigm.

Adoption of the aforementioned novel formal approach will drastically improve the quality of software on which our daily lives rely. Case studies from the IoT domain will drive the research and demonstrate the effectiveness of the approach. In particular, building on experience in developing and analysing IoT and railway systems in EU projects [R2], and exploiting background in requirements elicitation from domain experts, FMT will empirically investigate the effectiveness of the proposed techniques and methodologies on application scenarios related to the smart maintenance and monitoring of the Italian socio-ecosystem with its inherent variability. This will be carried out by means of case-study research, with structured and rigorous protocols [R3], in order to ensure empirical evidence and provide the involved practitioners with a clear view of the potential of the approach, while revealing practical challenges (e.g., tool usability, technical skills required, etc.) that can help to fine-tune the project's output. Furthermore, FMT will perform evaluations of human factors (e.g., usability, comprehensibility, etc.) in IoT software design and implementation.

T-LADIES will run until May 2025 and is coordinated by Walter Cazzola from the University of Milan. Other partners are the University of Catania, the University of Genoa, the University of Modena and Reggio Emilia, and ISTI-CNR.

Link:
[L1] https://t-ladies.di.unimi.it/

References:

[1] D. Basile, M.H. ter Beek, “Contract Automata Library”, Sci. Comput. Program, 2022. http://doi.org/10.1016/j.scico.2022.102841
[2] A. Ferrari, F. Mazzanti, D. Basile, M.H. ter Beek, “Systematic Evaluation and Usability Analysis of Formal Methods Tools for Railway Signaling System Design”, IEEE Trans. Software Eng., 2021. https://doi.org/10.1109/TSE.2021.3124677
[3] M.H. ter Beek, A. Ferrari, “Empirical Formal Methods: Guidelines for Performing Empirical Studies on Formal Methods”, Software, 2022. https://doi.org/10.3390/software1040017

Please contact:

Maurice ter Beek or Alessio Ferrari
ISTI-CNR, Italy
maurice.terbeek@isti.cnr.it, alessio.ferrari@isti.cnr.it

Heritage Annotator: Documenting, Browsing and Exploring Culture Heritage Data

by Kostas Petrakis, Manos Paterakis and Dimitris Angelakis (ICS-FORTH)

Heritage Annotator is a tool to support the creation and management of semantically structured annotations on digital representations of cultural heritage objects for the documentation of the acquired data and the results of analytical examinations, as well as of conservation condition reports or remedial works performed on the objects.

A vast area of research in Heritage Sciences concerns the documentation of analytical examinations and conservation interventions on heritage objects carried out by different actors. Archaeologists, conservators, engineers, material scientists, curators, and restorers of cultural assets are constantly enriching the knowledge and information about heritage artefacts. As expected, this knowledge consists of highly heterogeneous data produced by different procedures. Current practices make use of spreadsheets or text files to organise the information. Although these forms offer data analysis and scholarly interpretation, they pose problems including: i) difficulty for collaborative but controlled documentation by a large number of users, ii) an inability to represent the details from which the documented relations are inferred, iii) difficulty in extending the underlying data structures as well as in combining and integrating data from multiple and diverse sources and procedures, and iv) limitations in reusing the data beyond the context of a particular research activity.

The need to store and identify this information under a common denominator is more than obvious. The basic idea for the Heritage Annotator is to meet this need by allowing users to directly annotate images of the object, whether it is a painting, a sculpture, a building, or a work of art in general. This approach creates a clear visual overview of the procedures that

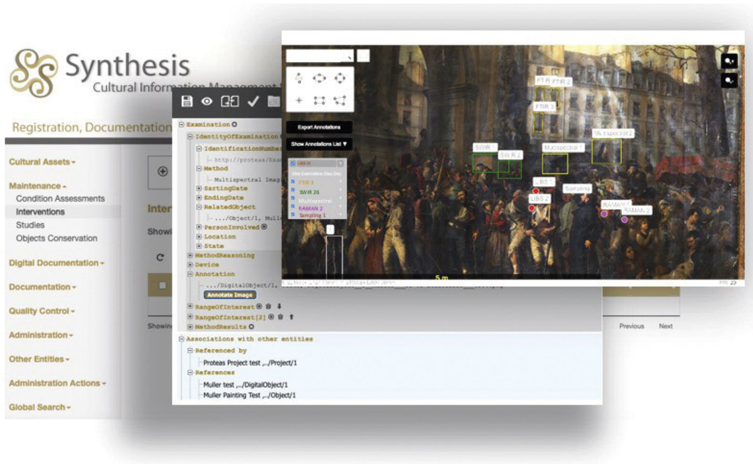


Figure 1: An example of how Heritage Annotator tool can be accessed by the Synthesis documentation system.

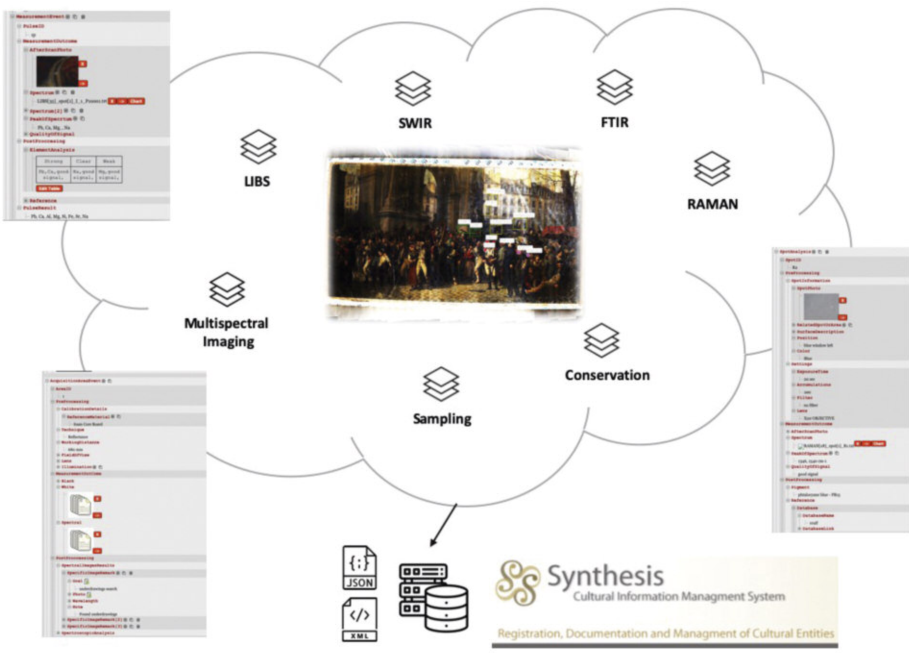


Figure 2: Overview of how the data are stored.

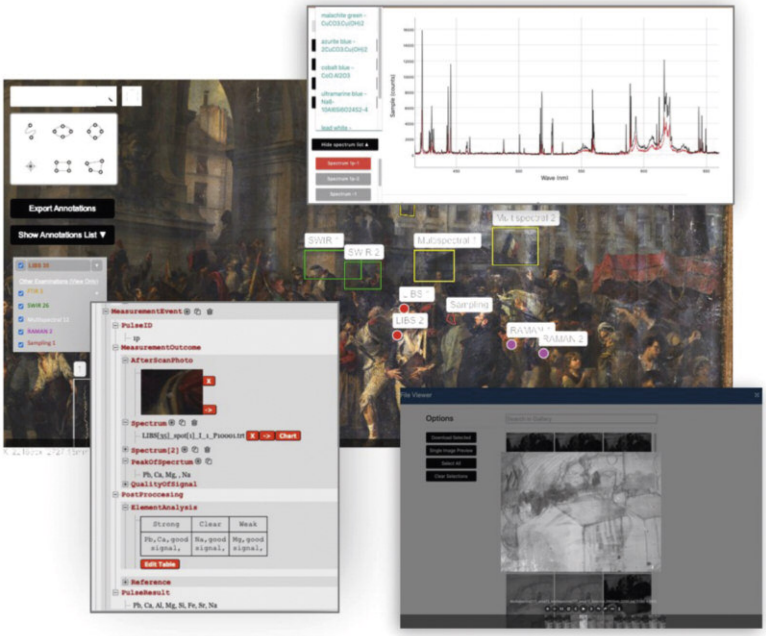


Figure 3: Heritage Annotator interface and functionalities.

have been followed and the location from which this information was derived. The annotation data structure makes use of existing standards for information documentation and publication (CIDOC-CRM) [L1], focusing on semantic interoperability and the production of searchable data of high value and long-term validity. These data are stored and managed on a common database, which is provided by the Synthesis Documentation System of FORTH [1] (see Figure 1).

In this article, we describe the process of documenting on the same digital representation of a heritage object, and analytical techniques that provide different information and present different degrees of invasiveness, which implies a hierarchical protocol for their sequential application. Additionally, to allow the user to preview and annotate the areas of examination with high accuracy, regardless of the scale of operation, a function of dynamic zoom is available. As an example, MultiSpectral Imaging on an area, microRAMAN at micro scale, and macroLIBS (Laser Induced Breakdown Spectroscopy) inducing a macro-spot of ablation, SWIR (Short-wave infrared) and FTIR (Fourier-transform infrared spectroscopy) examinations are presented (see Figure 2).

All annotations are linked to the same digital image of the heritage object, creating rich, structured knowledge, which helps scientists from different fields to combine their data and cooperate on curation and conservation. This means that each examination has its own different sematic schema for the annotations. Data are stored on a different record for every analytical examination into the same database and the tool gives the ability to link up annotations with scientific examination records. Advanced search is available on specific annotation parameters and results, and the tool provides a mechanism for exporting/importing examination records in xml format. Synthesis documentation system provides a detailed overview of all cultural heritage objects and all the linked activities that have been done in the past (see Figure 3).

Heritage Annotator’s basic functionality is that it allows the user to create/edit and delete annotations. It also provides the ability to view all the related annotations on the same object supporting text search on all annotations of an object. Additionally, the tool supports spatial search for annotations in a specific area and makes use of multiple image layers (e.g., a different photo of the same object after a long period or intervention). Heritage Annotator is also capable of deep zooming, can support large image files by using tiles, and handles chart visualisation for spectral data. Finally, the tool supports chart comparison with pigment refence databases facilitating pigment comparisons for those performing the examination. Further information about this tool can be found here [L2].

Acknowledgment
The financial support of this work has been provided by the project “PROTEAS” (Advanced system for collection and management of analytical data for documentation and conservation of large-scale paintings in an open laboratory, MIS 5069984), which are respectively implemented under the “Action for the Strategic Development on the Research and Technological Sector” and “RESEARCH-INNOVATE-CREATE II” and are funded by the Operational Program ¨Competitiveness, Entrepreneurship and Innovation¨ (NSRF 2014-2020) and co-financed by Greece and the European Union (European Regional Development Fund).

Links:
[L1] https://www.cidoc-crm.org/
[L2] https://demos.isl.ics.forth.gr/Heritage-Annotator/

Reference:
[1] P. Fafalios et al., “Towards Semantic Interoperability in Historical Research: Documenting Research Data and Knowledge with Synthesis”, in Proc. of 20th Int. Semantic Web Conf. (ISWC 2021), 2021.

Please contact:
Kostas Petrakis, ICS-FORTH, Greece
+30 2810 391488, cpetrakis@ics.forth.gr

ERCIM "Alain Bensoussan" Fellowship Programme

The ERCIM PhD Fellowship Programme has been established as one of the premier activities of ERCIM. The programme is open to young researchers from all over the world. It focuses on a broad range of fields in Computer Science and Applied Mathematics.

The fellowship scheme also helps young scientists to improve their knowledge of European research structures and networks and to gain more insight into the working conditions of leading European research institutions. The fellowships are of 12 months duration (with a possible extension), spent in one of the ERCIM member institutes. Fellows can apply for second year in a different institute.

Why to apply for an ERCIM Fellowship?

The Fellowship Programme enables bright young scientists to work on a challenging problem as fellows of leading European research centers.An ERCIM fellowship helps widen and intensify the network of personal relations among scientists.

The programme offers the opportunity to ERCIM fellows:
- to work with internationally recognized experts;
- to improve knowledge about European research structures and networks;
- to become familiarized with working conditions in European research centres;
- to promote cross-fertilization and cooperation, through the fellowships, between research groups working in similar areas in different laboratories.

Where are the fellows hosted?

Only ERCIM members can host fellows. When an ERCIM member is a consortium the hosting institute might be any of the consortium's members. When an ERCIM Member is a funding organisation, the hosting institute might be any of their affiliates. Fellowships are proposed according to the needs of the member institutes and the available funding.

Conditions

Candidates must:
- have obtained a PhD degree during the last eight years (prior to the application year deadline) or be in the last year of the thesis work with an outstanding academic record. Before starting the grant, a proof of the PhD degree will be requested. ;
- be fluent in English.

The fellows are appointed either by a stipend (an agreement for a research training programme) or a working contract. The type of contract and the monthly allowance/salary depends on the hosting institute.

ERCIM encourages both researchers from academic institutions and scientists working in industry to apply.

Equal Opportunities

ERCIM is committed to ensuring equal opportunities and promoting diversity. People seeking fellowship within the ERCIM consortium are not discriminated against because race, color, religion, gender, national origin, age, marital status or disability.

Application deadlines

Deadlines for applications are currently 30 April and 30 September each year.

Since its inception in 1991, over 750 fellows have passed through the programme. In 2021, 26 young scientists commenced an ERCIM PhD fellowship and 54 fellows have been hosted during the year. Since 2005, the Fellowship Programme is named in honour of Alain Bensoussan, former president of Inria, one of the three ERCIM founding institutes.

http://fellowship.ercim.eu

"The Alain Bensoussan fellowship at NTNU has been a great opportunity for me to establish a research career by providing excellent research conditions: freedom to do in-depth research, trainings for individual skill development, and a great network for interdisciplinary collaborations.

Philipp TERHÖST
Former ERCIM Fellow

"ERCIM fellowship at VTT was an amazing experience. VTT provided me with excellent research environment and facility and it helped me to get introduced to experts in my research field and work as I intended to. It also helped me to collaborate with many other researchers of other ERCIM institutes through the research exchange programme and seminars. Overall, it was a great learning experience and helped me build up a scientific career.

Firdausa AHMED
Former ERCIM Fellow

SCHLOSS DAGSTUHL
Leibniz-Zentrum für Informatik

Call for Proposals

Dagstuhl Seminars and Perspectives Workshops

Schloss Dagstuhl – Leibniz-Zentrum für Informatik is accepting proposals for scientific seminars/workshops in all areas of computer science, in particular also in connection with other fields.

If accepted, the event will be hosted in the seclusion of Dagstuhl's well known, own, dedicated facilities in Wadern on the western fringe of Germany. Moreover, the Dagstuhl office will assume most of the organisational/ administrative work, and the Dagstuhl scientific staff will support the organizers in preparing, running, and documenting the event. Thanks to subsidies the costs are very low for participants.

Dagstuhl events are typically proposed by a group of three to four outstanding researchers of different affiliations. This organizer team should represent a range of research communities and reflect Dagstuhl's international orientation. More information, in particular details about event form and setup, as well as the proposal form and the proposing process, can be found on

https://www.dagstuhl.de/dsproposal

Schloss Dagstuhl – Leibniz-Zentrum für Informatik is funded by the German federal and state government. It pursues a mission of furthering world class research in computer science by facilitating communication and interaction between researchers.

Important Dates
- *Next submission period:* October 15 to November 1, 2022
- *Seminar dates:* Between November 2023 and August 2024.

Inria

Throughout the year, Inria welcomes new employees to its research teams and departments, whether through competitions, mobility within the public service, contractual agreements or internship proposals

https://jobs.inria.fr/public/classic/en/offres

ERCIM
European Research Consortium for Informatics and Mathematics

Horizon Europe Project Management

A European project can be a richly rewarding tool for pushing your research or innovation activities to the state-of-the-art and beyond. Through ERCIM, our member institutes have participated in more than 100 projects funded by the European Commission in the ICT domain, by carrying out joint research activities while the ERCIM Office successfully manages the complexity of the project administration, finances and outreach.

Horizon Europe:How can you get involved?
The ERCIM Office has recognized expertise in a full range of services, including:
- Identification of funding opportunities
- Recruitment of project partners (within ERCIM and through ournetworks)
- Proposal writing and project negotiation
- Contractual and consortium management
- Communications and systems support
- Organization of attractive events, from team meetings to large-scale workshops and conferences
- Support for the dissemination of results.

Please contact:
Peter Kunz, ERCIM Office
peter.kunz@ercim.eu

ERCIM – the European Research Consortium for Informatics and Mathematics is an organisation dedicated to the advancement of European research and development in information technology and applied mathematics. Its member institutions aim to foster collaborative work within the European research community and to increase co-operation with European industry.

ERCIM is the European Host of the World Wide Web Consortium.

Consiglio Nazionale delle Ricerche
Area della Ricerca CNR di Pisa
Via G. Moruzzi 1, 56124 Pisa, Italy
www.iit.cnr.it

Norwegian University of Science and Technology
Faculty of Information Technology, Mathematics and Electrical Engineering, N 7491 Trondheim, Norway
http://www.ntnu.no/

Centrum Wiskunde & Informatica
Science Park 123,
NL-1098 XG Amsterdam, The Netherlands
www.cwi.nl

RISE SICS
Box 1263,
SE-164 29 Kista, Sweden
http://www.sics.se/

Fonds National de la Recherche
6, rue Antoine de Saint-Exupéry, B.P. 1777
L-1017 Luxembourg-Kirchberg
www.fnr.lu

SBA Research gGmbH
Floragasse 7, 1040 Wien, Austria
www.sba-research.org/

Foundation for Research and Technology – Hellas
Institute of Computer Science
P.O. Box 1385, GR-71110 Heraklion, Crete, Greece
www.ics.forth.gr

SIMULA
PO Box 134
1325 Lysaker, Norway
www.simula.no

Fraunhofer ICT Group
Anna-Louisa-Karsch-Str. 2
10178 Berlin, Germany
www.iuk.fraunhofer.de

Eötvös Loránd Research Network
Számítástechnikai és Automatizálási Kutató Intézet
P.O. Box 63, H-1518 Budapest, Hungary
www.sztaki.hu/

INESC
c/o INESC Porto, Campus da FEUP,
Rua Dr. Roberto Frias, nº 378,
4200-465 Porto, Portugal
www.inesc.pt

University of Cyprus
P.O. Box 20537
1678 Nicosia, Cyprus
www.cs.ucy.ac.cy/

Institut National de Recherche en Informatique
et en Automatique
B.P. 105, F-78153 Le Chesnay, France
www.inria.fr

Universty of Warsaw
Faculty of Mathematics, Informatics and Mechanics
Banacha 2, 02-097 Warsaw, Poland
www.mimuw.edu.pl/

I.S.I. – Industrial Systems Institute
Patras Science Park building
Platani, Patras, Greece, GR-26504
www.isi.gr

VTT Technical Research Centre of Finland Ltd
PO Box 1000
FIN-02044 VTT, Finland
www.vttresearch.com