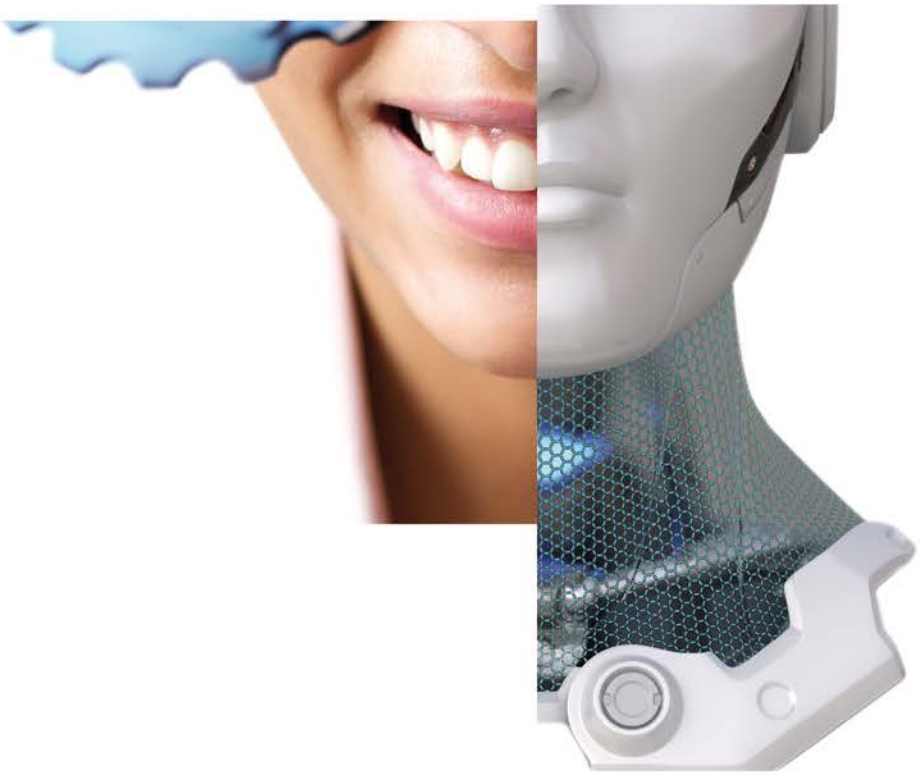
2020

TUM IEAI

TUM Institute for Ethics in Artificial Intelligence



ANNUAL REPORT

TUM IEAI

2020

Despite its difficulties, 2020 has been rewarding.

TUM

Institute for Ethics in Artificial Intelligence

ANNUAL REPORT

2020

ieai.mcts.tum.de

Index

Director's Message

Dear Friends & Associates

AI is here to stay and will continue to transform society substantially. The most pressing question is how this transformation will take place and what the repercussions will be. The Institute for Ethics in Artificial Intelligence (IEAI) seeks to be a leader in this process by exploring the ethical issues related to the development, use and impact of AI-enabled technologies. Founded in 2019, the IEAI celebrated its first anniversary at the end of 2020.

This first year was both exciting and intense. We have come a long way under unprecedented times. Our ability to react rapidly to unforeseen events, such as the COVID-19 pandemic was tested in practice. This health crisis demanded flexibility and rapid response and I am very proud to say that we have accomplished great things throughout this challenging year.

We have significantly grown our human capital. The IEAI's amazing international team now consists of 12 committed team members, 15 remarkable project Principle Investigators and more than 25 highly skilled researchers, experts and doctoral students. We plan to continue to expand in 2021 with new projects and researchers. From the very beginning, our goal was to create a research institute that brings together people, who support each other in addressing real-world challenges and far-reaching future questions through interdisciplinary research.

As the research project reports will display, we are constantly and effectively working towards achieving this goal. This year, **we have established interdisciplinary research groups and research clusters in a number of important areas. Seven clusters represent the IEAI's current key research areas,** with 14 individual projects conducted within these clusters. In addition to the more than €2.3 million we awarded to our eight initial projects, as an immediate response to COVID-19 crisis, we allocated over €400,000 in 2020 to projects aimed at research on AI and its use in health-related crises. Through these projects, we have created a research institute that focuses on practical and action-oriented outcomes and aims to develop tangible frameworks, methodologies and algorithmic tools. Moreover, we initiated our IEAI Research Brief series, which published its first six issues this year.

We have also expanded our partnerships around the world. With the launch of the Global AI Ethics Consortium (GAIEC) and the Responsible AI Network – Africa (RAIN – Africa), we are now working with AI ethics experts on six continents. We have also started working with Munich based companies, such as Celonis, on student-driven research. These initiatives have strengthened our goal of becoming a platform for debate, discussion and cooperation between all AI stakeholders. A goal we will continue to build upon, through events and research activities in 2021.

The COVID-19 pandemic has fundamentally altered how we plan, organize and host meetings and events. However, these trends also provided us with an opportunity to allow our speakers and workshop members to reach audiences all around the world. Between our launch in October 2019 and December 2020, the IEAI hosted **nine Speaker Series talks, four RAIN-Africa workshops, one international virtual forum,** as well as several other internal or co-hosted events.

AI is a powerful tool that has the ability to amplify current positive efforts, but also entrench and increase harm and discrimination if not used responsibly. This health crisis has made clear the need for a coordinated, dedicated data and technology infrastructure as well as an ecosystem for tackling the societal threats. It has also underlined the urgent need to discuss the ethical considerations in the use of AI and develop operational ethical frameworks.

We strongly believe that the technical and ethical problems are intertwined and the only way to tackle these issues is through interdisciplinary work that joins forces with all relevant stakeholders, including those from industry, civil society, government and academia. We strongly believe that working on a local, national or even regional basis on these matters is no longer enough. **AI is a global challenge, so is AI ethics.** From the ethical, as well as from the economic and societal side, AI ethics matter.

Despite its difficulties, 2020 has been rewarding. 2021 will bring us new challenges. However, it will also bring us new opportunities. We will see the first results of our research projects, launch the AI Ethics – Global Perspectives open access online course and have the chance to discuss ethics and governance of AI with distinguished international guest speakers in our Speaker Series, IEAI workshops and, of course, at The Responsible AI Forum **(TRAIF)** in Munich. Our major international conference will address the most relevant and pressing issues related to the responsible use of AI through shared stories, cutting-edge research and practical applications.

Last but not least, we will continue to build partnerships with other players inside, but also outside, of academia and strengthen our relationships with our valuable network partners, including the AI4People, GAIEC, ITU FG-AI4AD and RAIN-Africa. Through this, we will be able to expand our work and relationships with other major institutions around the world.

During the last year, the IEAI has been consolidated to a large degree. It now has a strong "heart" through its projects and a distinct "soul" through its people. Many challenges related to AI ethics still lie ahead and the IEAI is making a clear and concerted effort to put these topics higher on the AI agenda.

Without ethics, AI cannot fly. ●

Yours,

Christoph Lütge
Director

Our institute

TUM has long been a driving force in researching the mutual interactions of science, technology and society. Founded in 2019, the Institute for Ethics in Artificial Intelligence (IEAI) follows TUM's strategy for "Human-Centered Engineering" and the university's bold creation of the Munich Center for Technology in Society (MCTS) in 2012, whose mission is to better understand and reflexively shape the multiple interactions between science, technology and society.

Society-relevant innovation cycles
are not possible
without the ethical, legal
and political overall view.

[Prof. Thomas F. Hofmann, TUM President]

Our Mission.

The increasing involvement of technology in almost every aspect of society has already brought about a fundamental change in an unprecedented scope of areas. It goes without question that AI already has and will continue to transform society substantially within the course of our lifetimes. The most pressing question now is how this societal transformation will take place, and what the repercussions will be. The IEAI poses questions such as, "what should be possible in AI?" and "how can we ensure that as many people as possible benefit from AI's rewards?"

IEAI's foremost priority is the generation of global, egalitarian and interdisciplinary guidelines for the ethical development and implementation of AI.

Our Contribution to the AI Ethics Field.

- Support and carry out outstanding multidisciplinary research.
- Provide a platform for meaningful cooperation between a wide range of important stakeholders.
- Serve as a resource for students and interested stakeholders for information on ethics and responsible AI.

Our Journey so Far ...

We have come a long way since our launch in October 2019 and have grown in terms of our human capital, research profile and activities. The heart of the IEAI is its projects. The soul of the IEAI is its people.

At TUM
we have the perfect prerequisites,
with cutting-edge research
not only in the field of Artificial Intelligence,
but also in the social sciences.
Our Institute is one of the places
where this extremely valuable combination
can be appropriately leveraged.

[Prof. Christoph Lütge]

During 2019-2020
the IEAI awarded
more than € 2.7 million
in funding for research
on AI Ethics.

IEAI established

7

new research clusters.

Research Clusters & Projects.

Start

1 AI and the future of work

- A Human Preference-aware Optimization System

2 AI, mobility and safety

- ANDRE – AutoNomous DRiving Ethics

3 AI, choice and autonomy

- Consumer Perception and Acceptance of an AI-enabled Recommender System

4 AI in medicine and healthcare

- Building Strategic Partnerships to Understand Ethics and the Use of AI to Manage Health-related Crises
- ETHAN – Ethical AI for Pandemic Management
- METHAD - Toward a *ME*dical *ETH*ical *AD*visor System for Ethical Decisions
- Public Trust in AI and the Ethical Implications: A Comparative Study of Governmental Use of AI during the COVID-19 Pandemic
- The Ethics and Practice of AI Localism at a Time of COVID-19 and Beyond
- Rule of Law, Legitimacy and Effective COVID-19 Control Technologies

5 AI and online behavior

- Online Firestorms and Resentment Propagation on Social Media: Dynamics, Predictability and Mitigation
- Online-offline Spillovers – Potential Real-world Implications of Online Manipulation
- Personalized AI-based Interventions Against Online Norm Violations: Behavioral Effects and Ethical Implications

6 AI, governance and regulation

- TrustMLRegulation – Managing Trust and Distrust in Machine Learning with Meaningful Regulation

7 AI and sustainability

- Artificial Intelligence for Earth Observation (AI4EO): Reasoning, Uncertainties, Ethics and Beyond

2
new networks founded

More than
55
valuable individuals affiliated with the IEAI

6
new partners and projects funders

7
established research clusters

6
IEAI produced policy focused research briefs

8
reflections on AI-Q&A series with some of our distinguished guest speakers

18
public events hosted

13
published and submitted books and articles

14
research projects across

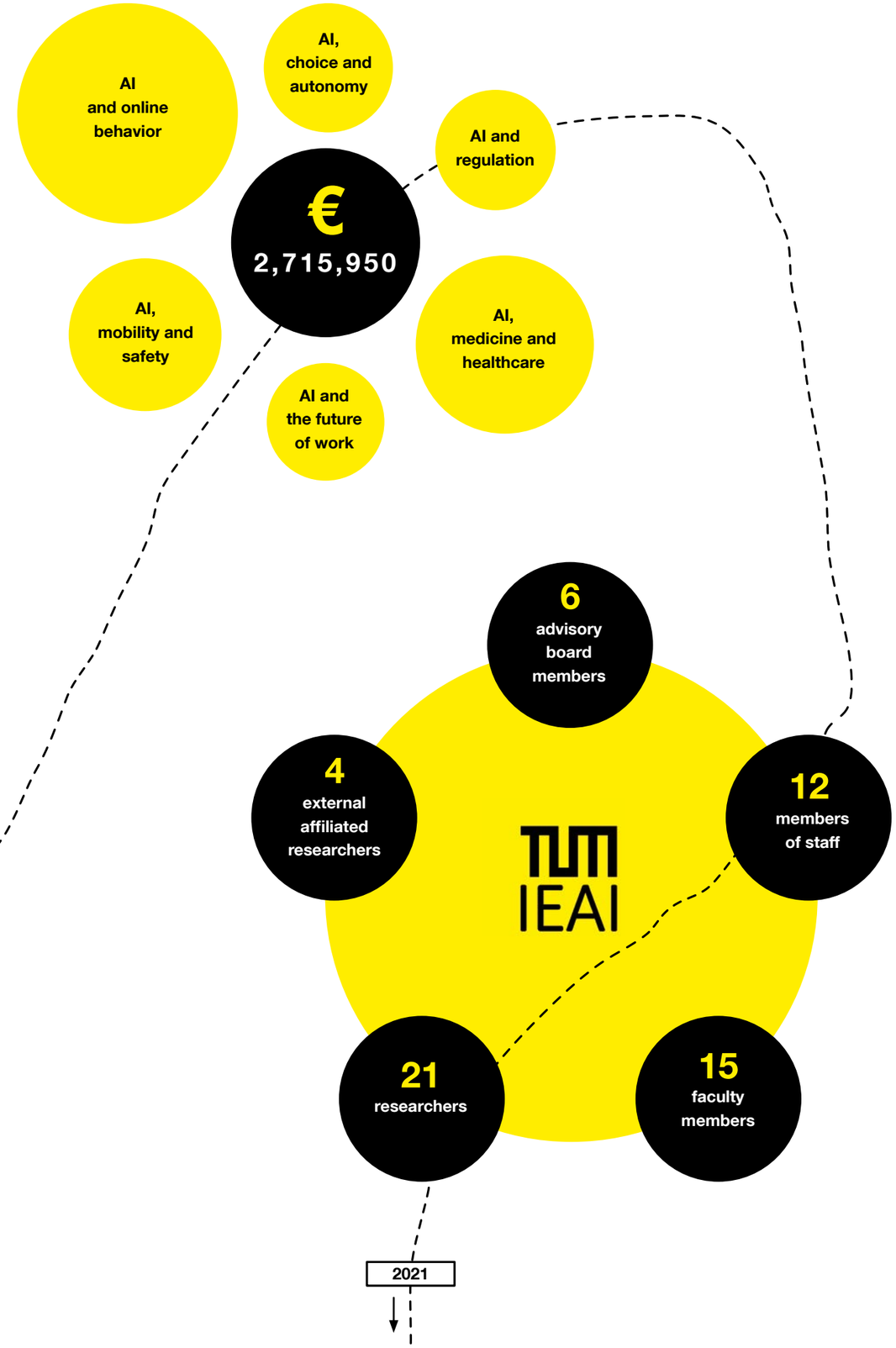
11
schools at TUM and

7
external academic partners fom across the globe

IEAI speakers at more than
59
German, European and international events

1,186
event attendees

More than
3,000
social media followers



Structure and Organization
Advisory Board

The IAEI's activities are guided by an international advisory board composed of expert representatives from academia, industry and civil society. This group evaluates the work and progress of the IEAI and makes recommendations on its future direction.

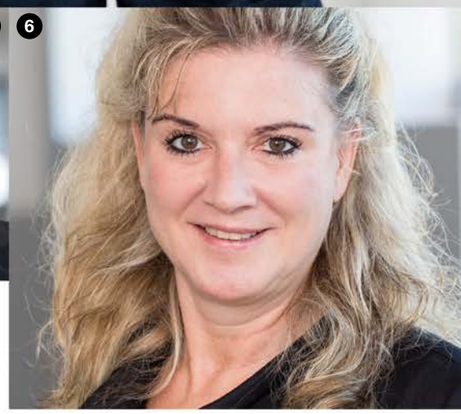
Its members are part of the Institute's valuable human capital and have decades of experience in their respective fields. They represent a broader audience with which the IEAI seeks to have a strong connection. The members are influential members of their community with a big picture understanding of the importance of AI ethics. Their advice provides a third-party perspective, adds value by exposing the IEAI team to new thinking and widens the networks of contacts and opportunities available to the IEAI.

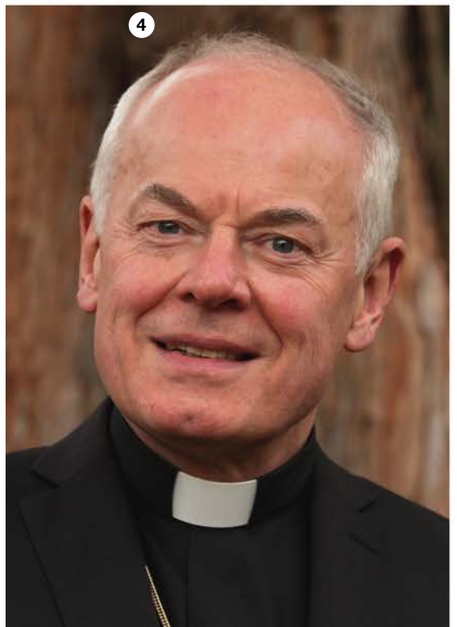
1
2

In my view,
academic institutions with robust ethical oversight
should lead the way in terms of scientific
and technological progress. This is necessary
in order to ensure that developments of AI
are not hijacked by private interests
of purely economic nature, but advancements
are made with public interest in mind,
and that the principles of research ethics
and human rights are adhered to.

Being a member of the IEAI advisory board
is a rare and exciting opportunity to obtain
first-hand insights into the newest trends
in AI research.
My personal motivation to become involved
was related to the ambition of giving human rights
a more pronounced role in the debates
about the direction of AI developments.

[Zuzanna Warso, Advisory Board Member at the TUM IEAI]





1 **Prof. Dr.-Ing. Dr.-Ing. E.h. Dr. h.c. Dieter Spath**
President of acatech and Director of Fraunhofer IAO and IAT, University of Stuttgart

2 **Zuzanna Warso**
Helsinki Foundation for Human Rights, Warsaw

3 **Prof. Dr. Christoph Meinel**
Director and CEO of the Hasso-Plattner-Institute for Digital Engineering at the University of Potsdam

4 **Auxiliary Bishop Anton Losinger**
Episcopal Vicar („Bischofsvikar") for Bioethics and Social Policy, Augsburg

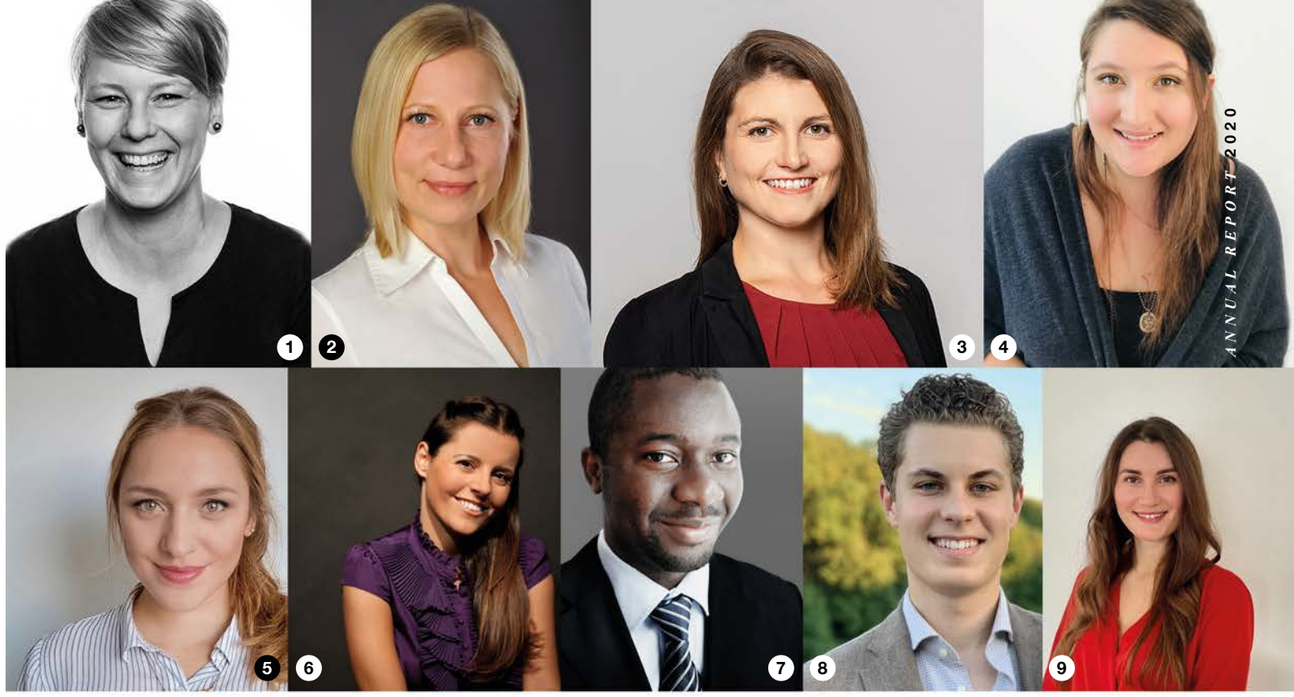
5 **Hannes Schwaderer**
President of Initiative D21 e. V. and CEO of Intel Germany

6 **Prof. Dr. Petra Ahrweiler**
Chair for Technology and Innovation Sociology, Johannes Gutenberg University Mainz

Structure and Organization
Staff

At the IEAI,
diversity and global mindedness
are integral to our DNA.

As with any working environment, our personnel are as important as the IEAI itself. Our employees are valuable assets of the Institute and the key to its success.
We firmly believe that a diverse and inclusive workforce is one of the best ways to ensure the development of new ideas. Thus, we are proud of our international team that consists of committed members from Africa, Europe and North America. From our office managers, to our researchers, to our student assistants, our high-level of productivity and inspired results is due to the commitment of our team. We asked our team about the values they recognized as most important in their enjoyment working at the institute. Here are their thoughts:



1 **Christina Daman**
Office, Program & Event Manager

2 **Manuela Fuchs**
Office, Program & Event Manager

3 **Elizabeth Weschenfelder**
Office Manager

4 **Auxane Boch**
Researcher

5 **Ellen Hohma**
Researcher

6 **Anastasia Aritzi**
Communications Consultant

7 **Laud Ammah**
Student Assistant

8 **Immanuel Klein**
Student Assistant

9 **Laura Lucaj**
Student Assistant

What I appreciate about the work at the IEAI,
is the interdisciplinary nature of the projects.
Ethics in AI is clearly a field of research
that incorporates several disciplines.
The IEAI brings people from different backgrounds
together, enabling integrative research
with a high level of expertise
in every field.

[Ellen Hohma, Research Associate]

When I think about our team at the IEAI,
the first adjectives that come to mind are multicultural,
diverse and interdisciplinary. The variety of profiles,
professional and personal, allow us to evolve
in an enriching and stimulating environment.
Our work on the fascinating topic of AI ethics benefits
from the support and advice we provide to each other,
leading to even more relevant outcomes.

[Auxane Boch, Research Associate]

The team was asked which aspects of the workplace at the IEAI were most important to them. The word cloud below displays the importance of our work environment attributes.

Their diverse cultural backgrounds along with their wide-ranging personal and professional experience, offer new perspectives to the team and at the same time allow our staff members to see their work and AI ethics-related challenges from various viewpoints and approaches. These multiple perspectives and personalities give rise to out-of-the-box thinking, boosting our problem-solving capacity and introducing new skills and approaches to our work. The goal is to further promote diversity within the IEAI: from our team members, principle investigators and researchers to our guest speakers, panel participants and external partners. Diversity and inclusion is not only an opportunity to incorporate the different viewpoints that are extremely necessary for contributing to ethical and responsible AI, but also provides a path towards increased equality.

In order to support the effective and safe advancement of AI technologies that will benefit all parts of society, ethical concerns associated with the use of AI need to be adequately addressed.

IEAI's inter-, multi-, and transdisciplinary research approach seeks to identify these challenges, deliver practical and action-oriented outcomes and develop tangible frameworks, methodologies and algorithmic tools in order to provide AI developers and practitioners with a set of ethical best practices.

What makes IEAI's research approach so distinctive is that the researchers work in interdisciplinary pairs: researchers from the technical side (e.g. computer scientists or engineers) working together with researchers from humanities, social sciences or law. This research approach enables us to truly and comprehensively address a growing group of ethical challenges arising at the interface of technology and human values.

Caitlin Corrigan, PhD
Scientific Coordinator
IEAI

Independent academic research
has a vital role to play in designing
AI-enabled technologies that have the potential
to truly make a positive impact on society.
University-based research has access
to the expertise and time horizons
to produce meaningful and high-quality results
to inform the complex challenges
we face in the field of AI.

[Caitlin Corrigan, PhD – Scientific Coordinator IEAI]

At the IEAI,
we aim to create an environment
where research
based on academic integrity
and interdisciplinary thinking can take place
and where relevant stakeholders
can engage with the research results and findings
in order to make a significant and informed impact
on AI policy and development.

Prof. Dr. Anna Baumert
TUM School of Education and
Max Planck Institute for Research
on Collective Goods

- **Personalized AI-based Interventions Against Online Norm Violations**

Prof. Dr. Christian Djeffal
Munich Center for
Technology in Society and
TUM School of Governance

- **Rule of Law, Legitimacy and Effective COVID-19 Control Technologies**

Prof. Jens Großklags, PhD
Department of Informatics, TUM

- **Personalized AI-based Interventions Against Online Norm Violations**

Prof. Dr.-Ing Johannes Fottner
Department of Mechanical
Engineering, TUM

- **A Human Preference-aware Optimization System**

Prof. Dr. Christoph Lütge
TUM School of Governance

- **ANDRE – AutoNomous DRiving Ethics**
- **Building Strategic Partnerships to Understand Ethics and the Use Of AI to Manage Health Related Crises**
- **Public Trust in AI and the Ethical Implications**

Prof. Dr. Jutta Roosen
TUM School of Management

- **Consumer Perception and Acceptance of an AI-enabled Recommender System**

Prof. Tim Büthe, PhD
TUM School of Governance

- **A Human Preference-aware Optimization System**

Prof. Dr. Georg Groh
Department of Informatics, TUM

- **Consumer Perception and Acceptance of an AI-enabled Recommender System**
- **ETHAN – Ethical AI for Pandemic Management**

Prof. Dr. Massimo Fornasier
Department of Mathematics, TUM

- **Online Firestorms and Resentment Propagation on Social Media**

Prof. Dr. Markus Lienkamp
Institute of Automotive
Technology, TUM

- **ANDRE – AutoNomous DRiving Ethics**

Prof. Dr. Jürgen Pfeffer
TUM School of Governance and
Department of Informatics, TUM

- **Online Firestorms and Resentment Propagation on Social Media**
- **Online-offline Spillovers**

PD Dr. Matthias Uhl
TUM School of Governance

- **Online-offline Spillovers**
- **The Ethics and Practice of AI Localism at a Time of COVID-19 and Beyond**

The IEAI's research addresses challenges such as ...

In addition to our director, scientific coordinator and staff, the IEAI research team consists of 15 distinguished TUM faculty members who function as project principle investigators.

They are supported by over 20 highly skilled doctoral and post-doctoral researchers that work closely together towards the development of operational ethical frameworks in the field of AI. Our new projects aimed at the use of AI in managing health crises, also involve Principle investigators from universities other than TUM. In these projects, we are working with eight external project leaders, as well as multiple applied partners.

In the following section, we outline our seven research clusters and the work of the specific projects within each cluster over the past year.

AI and the future of work.

How can we use AI-based data analysis to improve workplace efficiency without weakening employee autonomy?

A Human Preference-aware Optimization System

The project aims to explore the consequences of digitization for employees in logistics areas. Companies gather vast amounts of data about processes and, with modern machine learning techniques, it is easy to derive process optimizations. However, AI-based process analysis does not just hold the promise of efficiency gains. It also allows employers to assess individual employees behavior and productivity.

The research expands upon emerging ethical concerns regarding the loss of autonomy and rising stress levels in the workplace. In this project, the goal is to utilize AI in an ethical manner that respects fairness, transparency and explainability. The aim is to create guidelines and design rules for applying AI in human-centered processes, while considering the ethical perspective. The project is a collaboration between the Chair of Materials Handling, Material Flow and Logistics at the Department of Mechanical Engineering and the Chair for International Relations in the School of Governance.

In order to achieve these goals, the project conducted a preliminary study in 2020 using experts from the IEAI researcher groups and similar chairs in mechanical engineering to investigate requirements and risks for the project. To formally describe the sample workplaces, a morphological box based on literature approaches was developed from this study and identified appropriate key performance indicators (KPIs) for measuring efficiency. This resulted in five indicators for each workplace.

Next, in order to design worker preference models, basic personalities are combined with logistics job profiles. To this end, in 2020 a study was conducted in cooperation with a manufacturing company, where each participant fulfilled a work description sheet and wore a device to measure heart rate variability for 24 hours.

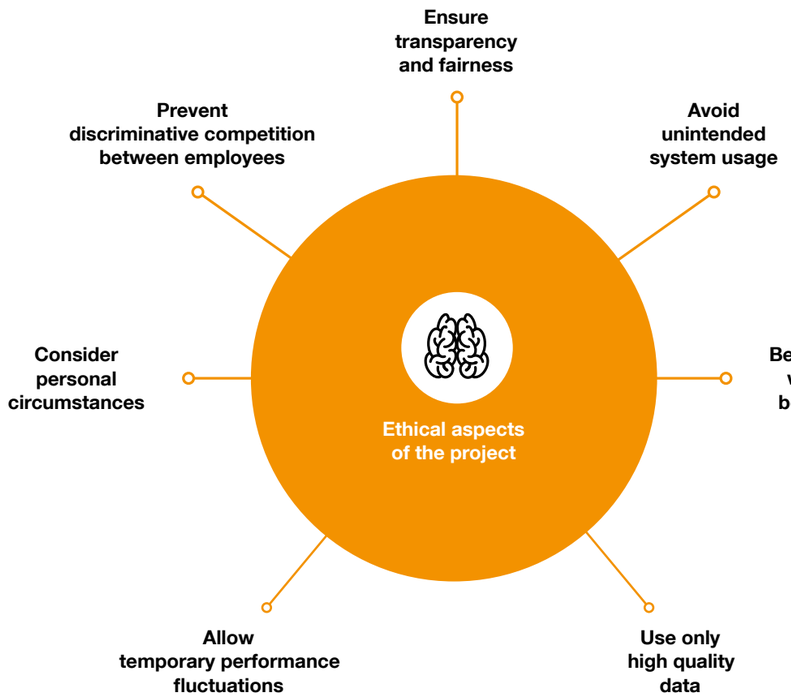
From this initial work, the team found that:

- transparency, fairness, data assessment, legal boundaries and technical safeguards are important requirements for an AI optimization system.
- for the optimization methodology, transparency, fairness, process improvement and dynamic adjustments were seen as requirements.
- first evaluations of smart wearables show that it is a very suitable way to measure job satisfaction for the project.
- many of the optimization problems encouraged in this research are similar to those existing in mathematics. However, initial analyses show that there are only a few approaches using AI and further research needed.

Plans for 2021

With the strong technical focus from 2020, the team plans to dive deeper into ethical considerations in 2021. Requirements, concerns and the state of knowledge in using AI in process optimization will be examined. The second focus will be on the data basis for AI optimization. With suitable hardware, initial studies on movement profiles and employee satisfaction with workplaces, will be conducted. From the collected logistics data, an AI model is to be built. With the model, different assumptions and data input can be investigated to gain insights for a handbook.

Ethical aspects of using
a decision-making AI in logistics



2020 Papers and Project Highlights

- Logistics process analysis and optimization methodology for a human preference-aware AI optimization system (Master's Thesis)
- Investigation to measure job satisfaction using fitness data from smart devices (Master's Thesis)
- Development of a personality test to improve the division of tasks in logistics (Semester's Thesis)

2020 Conferences

- The Responsible AI Forum (TRAIF) Preview 2020, November 2020

Principal Investigators

- Prof. Dr.-Ing. Johannes Fottner, Department of Mechanical Engineering, TUM
- Prof. Tim Büthe, PhD, TUM School of Governance

Researchers

- Dimitrij-Marian Holm, M.Sc., Department of Mechanical Engineering, TUM
- Charlotte Haid, M.Sc., Department of Mechanical Engineering, TUM
- Charlotte Unruh, TUM School of Governance

AI, mobility and safety.

How can we ensure the implementation of ethical behavior into autonomous vehicles?

ANDRE – AutoNomous DRiving Ethics

Autonomous vehicles (AVs) are expected to play a key role in future transportation systems. They will have a global impact that will change society as well as roadway and transportation system safety. Although AVs are assumed to improve overall road safety, they will not be able to eliminate all risks, which is flagged by accidents that have occured in connection with AVs over the past few years.

Therefore, a pressing question – that is also guiding the research of this project – is: How can a fair distribution of risk be realized in the trajectory planning of automated vehicles? The research project is a collaboration between the Institute of Automotive Technology and the Chair of Business Ethics at Technical University Munich. The interdisciplinary composition of the research team allows the creation of synergies and the combination of expert knowledge.The technical development within this project is the creation of software parts for autonomous vehicles. The developed software framework consists of four different modules, namely the trajectory planning, a risk assessment, trajectory prediction and a simulation environment. The focus here is on trajectory planning, which considers ethical aspects as part of a risk distribution. Finally, this software allows to test and empirically evaluate different ethical concepts.

Additionally, in terms of ethics related developments, the ANDRE-project aims to provide practical recommendations and guidelines for the automotive sector to serve as a checklist and moral compass on how to proceed and what to consider when developing AVs.

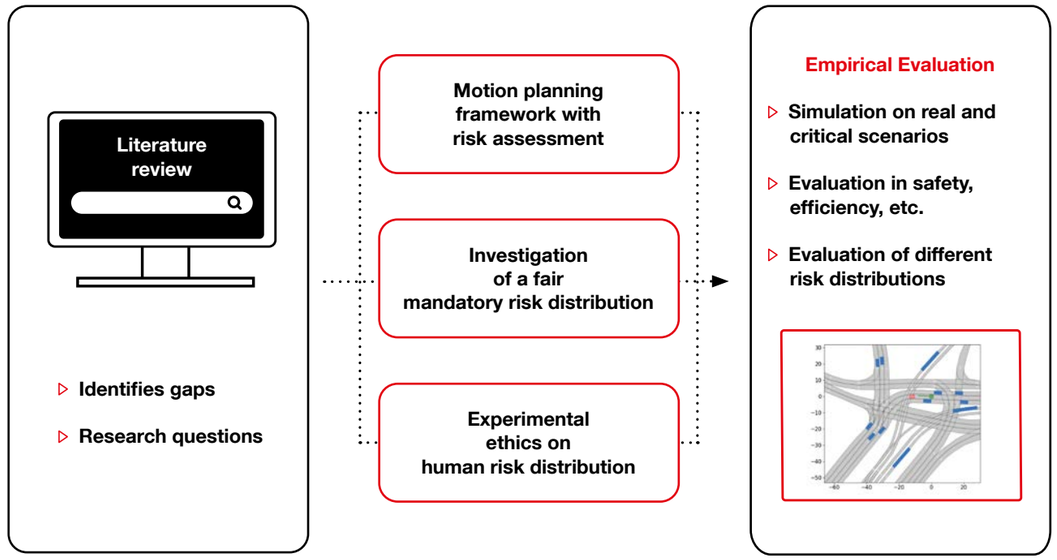
Preliminary findings from the project include:

- The discussion in the field of autonomous driving ethics needs to refocus: Instead of decision metrics for dilemma situations, as demanded by the trolley problem, a concept of responsible assessment and balancing of risks should be developed, especially for mundane traffic scenarios.
- Risk assessment can be integrated into the trajectory planning of AVs in the form of (a) an optimization problem and (b) validity checks (e.g., for maximum acceptable risk). A corresponding mathematical formulation of risk in the context of AVs is established within this project.
- First results show three maxims that are important for a responsible assessment and balancing of risks in the context of autonomous driving: Minimization of the overall risk, priority for the worst-off and equal treatment of people.
- Enforcing these maxims in practice would translate, amongst other aspects, into the prohibition to base balancing of risks on personal features of individuals, and instead base it on more objective factors that determine (a) collision probability and (b) estimated harm. Such factors can be the speed of particular traffic participants or the impact angle under which a collision would occur.

Plans for 2021

In 2021, the research team will finalize more publications, as well as conduct an experiment with the aim to investigate individuals' risk distribution preferences and detect potential thresholds of acceptable risk in traffic scenarios. Furthermore, the research team plans to organize a public roundtable to present and discuss the mid-term results of the ANDRE-project with key stakeholders in the automotive sector such as manufacturers, policymakers and consumers.

Project Overview



2020 Paper and Project Highlights

- AI4People: Ethical Guidelines for the Automotive Sector – Fundamental Requirements & Practical Recommendations for Industry and Policymakers, International Journal of Technoethics (C. Lütge, F. Poszler, A. Joaquin Acosta, D. Danks, G. Gottehrer, Lucian Mihet-Popa, A. Naseer)
- Programming Away Human Rights and Responsibilities? "The Moral Machine Experiment" and the Need for a More ‚Humane' AV Future, Nano Ethics (M. Kochupillai, C. Lütge, F. Poszler)
- Holistic Overview of the aApplicability of Traditional Ethical theories on Autonomous Vehicles. (under review)
- Autonomous Driving Ethics: From Trolley Problem to Ethics of Risk, Philosophy and Technology (under review) (M. Geißlinger, F. Poszler, J. Betz, C. Lütge, M. Lienkamp)

2020 Conferences

- 12th annual Forum for Humane Economic Order "Menschenwürdige Wirtschaftsordnung" at the Academy for Political Education in Tutzing, March 2020
- DiVA – Gesellschaftlicher Dialog zum vernetzten und automatisierten Fahren, February 2020
- Inaugural meeting of the ITU-T Focus Group on AI for Autonomous & Assisted Driving (FG-AI4AD)
- The Responsible AI Forum (TRAIF) Preview 2020, November 2020
- Virtual AI4People 2020 Summit

Principal Investigators

- Prof. Dr. Markus Lienkamp, Institute of Automotive Technology, TUM
- Prof. Dr. Christoph Lütge, TUM School of Governance

Researchers

- Maximilian Geißlinger, Institute of Automotive Technology, TUM
- Franziska Poszler, TUM School of Governance

AI, choice and autonomy.

How do consumers perceive AI-enabled technology aiding their consumption decisions and can these perceptions create barriers in the use of AI?

Consumer Perception and Acceptance of an AI-enabled Recommender System

Recommender systems (RS) are designed to ease and enhance consumer decision-making in complex online environments. Albeit devised to be convenient and ancillary, recommender systems may, in turn, influence and even dominate consumer's choices with available means of choice architecture. This raises concerns about consumer freedom of choice, which can be hindered by a recommender system without consumers being aware of being manipulated.

In a one year project, the team from the Chair of Marketing and Consumer Research collaborated with the team from the Department of Informatics to develop a survey that uses a mock-up of an online food ordering screen and an information treatment to determine factors influencing consumer acceptance of a recommender system that employs nudging as a method to steer consumer choice in the desired direction.

The goals of the study were to determine (i) which factors influence consumer perception of an AI-enabled recommender system, (ii) whether this perception differs when nudging is included in the system and (iii) how the perceived manipulation and privacy concerns impair the acceptance of a recommender system to facilitate decision-making.

A survey based on the Technology Acceptance Model (TAM) was constructed to determine consumer perception and acceptance of a recommender system.

The main findings of the study included that:

- Consumer acceptance is substantially determined by perceived system effectiveness and, to a lesser degree, by perceived ease of choice.
- Negative consumer concerns appeared to revolve around the information and algorithms used by an RS, but not the methods of choice architecture used to modify choice environments.
- Perceived manipulation and privacy concerns negatively affected users' acceptance of a recommender system.
- Tracking users' online activities without asking for permission and engaging in data trading were examples of conducts by e-commerce platforms that cause consumer privacy concerns.

From these findings, the team provided recommendations for AI-based e-commerce platforms to provide transparency in their handling of users' data and the reasons why certain items are recommended. Moreover, when business owners prompt certain products on AI-based e-commerce platforms, they should be more specific about the attributes of highlighted products. Instead of such descriptions as "popular choice" or "expert's recommendation", less ambiguous attribute descriptions like "healthier choice", "budget choice", or "most energy-efficient" can better facilitate consumers to navigate and select the products with features that meet their needs.

Plans for 2021

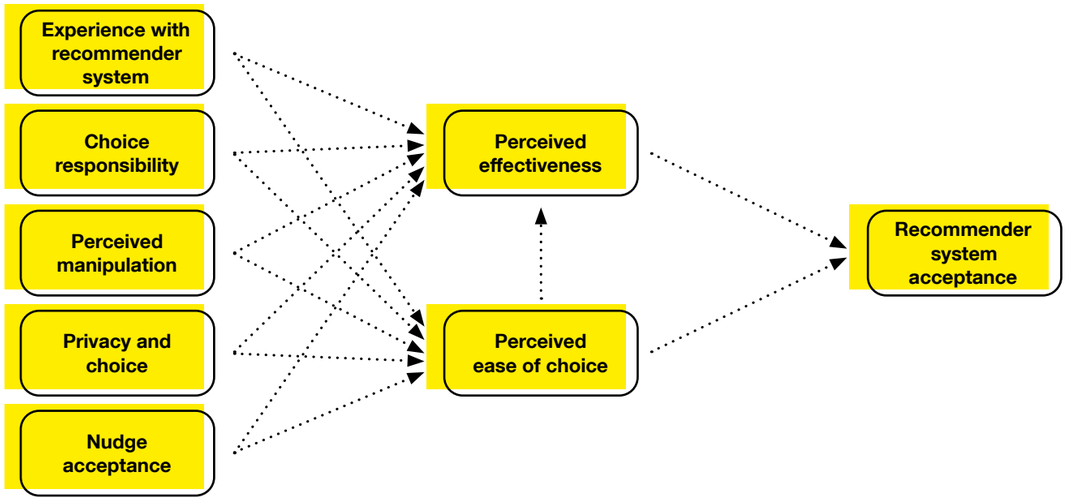
As the project winds down, the team is considering possible extensions of the research to include conducting more realistic and interactive experiments with a real recommender system or testing how people perceive manipulation and acceptance of nudges and RS vary with different types of nudges. They will also complete the publication of the manuscript in the Journal of Interactive Marketing.

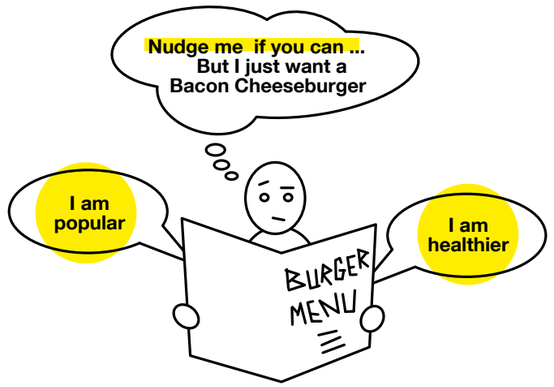
Principal Investigators

- Prof. Dr. Jutta Roosen, TUM School of Management
- Prof. Dr. Georg Groh, Department of Informatics, TUM

Researchers

- Dr. Irina Dolgopolova, TUM School of Management





2020 Papers, Projects and Achievements Highlights

- Best Poster Award 2020 (2nd place) at the TUM Research Fest
- Do Nudges Matter? Consumer Perception and Acceptance of Recommender Systems with Different Types of Nudges, (working paper) (I. Dolgopolova, B. Li, J. Roosen)
- Consumer Perception and Attitude Towards AI-Based Recommender Systems and the Use of Digital Nudging in Online Choice Environments (an example of an online food ordering screen) (Masters Thesis)
- Nudging Healthy Food Choices via a Recommendation System: Consumer Trust and Privacy Concerns (Masters Thesis)

2020 Conferences

- 3rd FAccTRec Workshop on Responsible Recommendation at RecSys 2020, September 2020
- The Responsible AI Forum (TRAIF) Preview 2020, November 2020
- Semester Seminar of the Department of Agricultural Economics and Rural Development at the University of Göttingen 2020, December 2020

AI in medicine and healthcare.

How can AI support physician and public health decision making in ethically responsible ways?

Building Strategic Partnerships to Understand Ethics and the Use of AI to Manage Health-related Crises

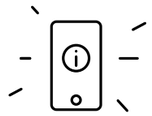
This yearlong project focuses on the use of AI-enabled technologies employed to monitor, respond to and manage individuals or populations deemed at risk. The research team will investigate two case studies, (1) the ethics of the use of AI in smart city technologies employed during crises (with partners at the Graduate School of Public Policy, University of Tokyo) and (2) ethics of the use of AI in adaptive data-driven health surveillance (with partners at the Dyson School of Design Engineering, Imperial College London).

In 2020, the case study of ethics of the use of AI in adaptive data-driven health surveillance was underway. Together with partners from the Imperial College London, our research team aims to develop ethical guidelines directed to policy-makers, to support the development and application of health surveillance technology to manage present and future health-related governance challenges. Personal health surveillance is defined as the continuous monitoring of health data and behavioral signals for wellbeing purposes. It allows, for example, caretakers and health professionals to detect changes in a patient's condition, or employers to monitor employees that might be exposed to pollutants, noise, vibration or other health hazards. AI technologies used in this context raise ethical, psychological and social consequence concerns.

Plans for 2021

This research aims to assist developers and policymakers in creating implementable guidelines that enable support at each step of the development process, with a major importance placed on aspects of fairness being integrated in the design. To better understand and target this issue, we will pilot stakeholder engagement in the form of surveys, interviews and behavioral experiments in 2021. A major output of this initial study will be the creation of a framework for ethical challenges in the different phases of the development of health surveillance tools. Anticipated further extensions of this product would aim to prototype the integration of the framework into new tools and designs.

These findings will be presented and further adapted in several workshops on ethics and AI-enabled technologies employed to monitor, respond to and manage individuals or populations at The Responsible AI Forum. Finally, to disseminate the results of this project, an educational strategy will be developed to communicate the ethical implications in health surveillance technology to developers and policymakers.



Principal Investigators

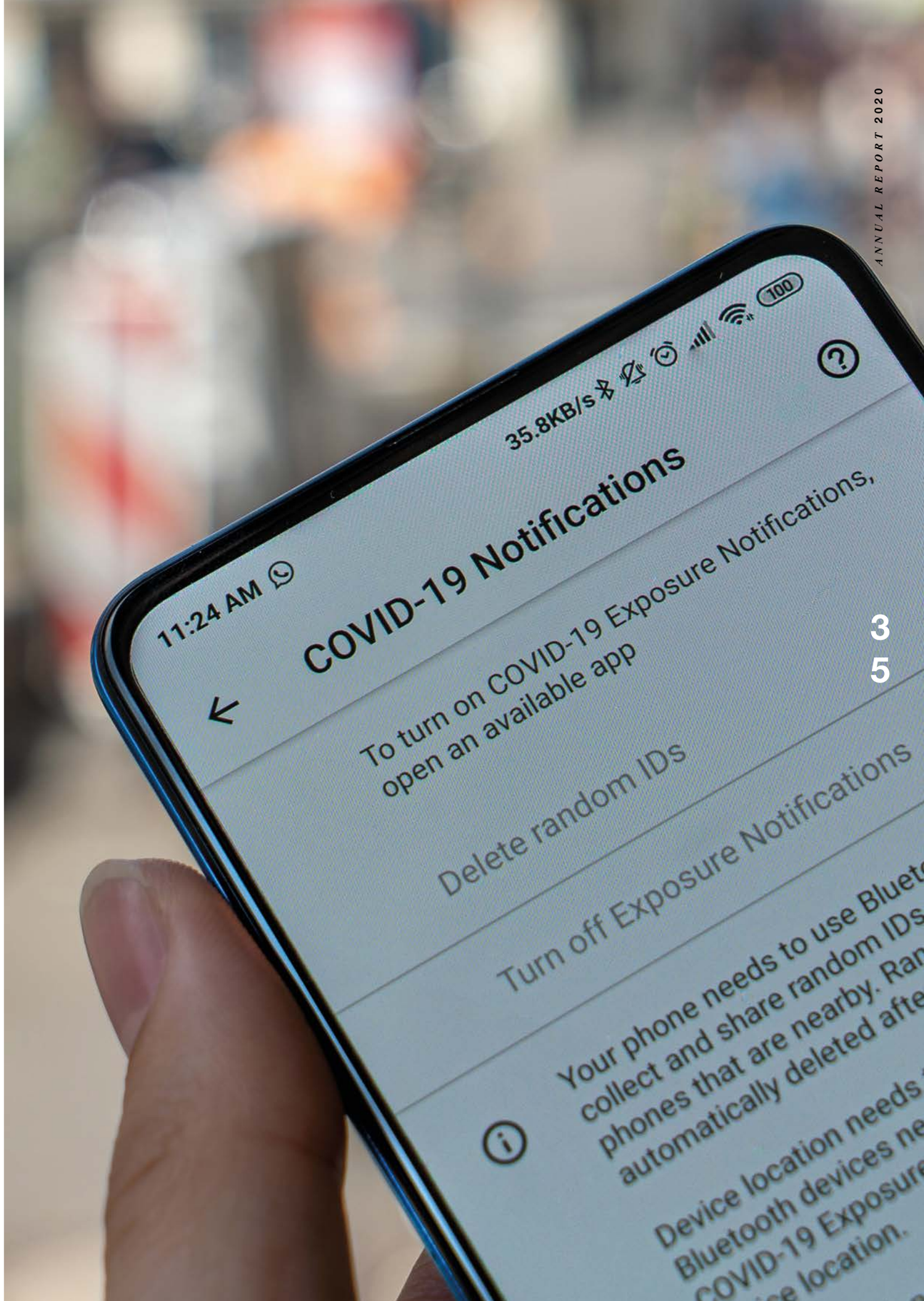
- Prof. Dr. Christoph Lütge, TUM School of Governance
- Prof. Rafael A. Calvo, PhD, Dyson School of Design Engineering, Imperial College London
- Prof. Kan Hiroshi Suzuki, Graduate School of Public Policy, University of Tokyo
- Prof. Dr. Ken Ito, Interfaculty Initiative in Information Studies, University of Tokyo
- Caitlin Corrigan, PhD, Institute for Ethics in Artificial Intelligence, TUM

Researchers

- Ellen Hohma, M.Sc., Institute for Ethics in Artificial Intelligence, TUM
- Ana Catarina Fontes, Institute for Ethics in Artificial Intelligence, TUM

Applied Partners

- UC Berkeley, Department of Mechanical Engineering
- Airport RFID Technology Agency (ARTA), Denso Wave Co.
- NPO Medical Governance Research Institute
- Bavarian State Tokyo Office



ETHAN –
Ethical AI for Pandemic Management

This new project looks at a scenario-based approach to the design and use of ethical AI models in managing a health pandemic and it is also part of the IEAI's efforts to promote research on the use of AI to manage public-health crises, while building upon our partnerships with the Global AI Ethics Consortium (GAIEC) members.

The collection and processing of such personal data raises many ethical questions, such as protection of private information, fairness, accountability and interpretability. This yearlong research project is a collaboration between the Department of Informatics at TUM, Stellenbosch University and Pralexis. It aims to develop an ethical and legal framework and possible machine learning (ML) strategies to manage health pandemics such as COVID-19 or HIV.

Key research questions undertaken are: What are the socio-ethical considerations when generating and processing personal data in health-related settings in terms of privacy, fairness and agency? What are the shortcomings of current approaches? What are the elements of a legal framework for AI-based epidemic management models and how could this be used in different jurisdictions in the world? Moreover, what are possible ML strategies that could be applied in managing major health crises based on an ethical and legal framework acceptable within a constitutional democracy?

The group is intensively working on a simulation environment, with one group aiming at simulating a realistic long-term-social network that is the basis for the regular contacts for disease spreading. The second group is working with the help of the tool mat_sim on simulating time series of human activities (spatial mobility and encounters) based on the long-term social network but also including random encounters. Using these data, a third group simulates the disease dynamics (e.g. the state changes for each human node) over the course of time.

The team uses extensive data from the South African Western Cape government that allows for fine-tuning of simulations. They also work with publicly available data from the Paris region to allow for computing realistic scenarios for large European urban environments. The applied partner, Pralexis, provides the student researchers with excellent computing resources. After completing the simulations, the group will work on a detailed planning stage to develope the data scenarios and the machine learning strategies. The scenarios involve various degrees of sampling from the ground truth data, corresponding to various degrees of privacy and public compliance as well as techniques such as differential privacy to allow for more sophisticated forms of privacy without severely harming the ML-based analysis.

Plans for 2021

The next step for the coming year is to develop ML-strategies. One area of thought aims at investigating risk prediction applications targeted for the individual citizen based on time, space, activities and parts of the social network that will mainly be facilitated by time-series-based predictive and analytical ML techniques. The other line of thought aims at a higher level of analysis targeted at government actors to decide upon pandemic control measures. Here, the group will investigate clustering approaches that will allow for an analysis of disease dynamics at a larger scale than the individual level, but more fine grained than mere descriptive statistics. In parallel, partners at the University of Stellenbosch are working on an extensive analysis of the legal aspects of AI-based pandemic control and a wide-ranging analysis of the ethical foundations for AI-based pandemic control.

Principle Investigators

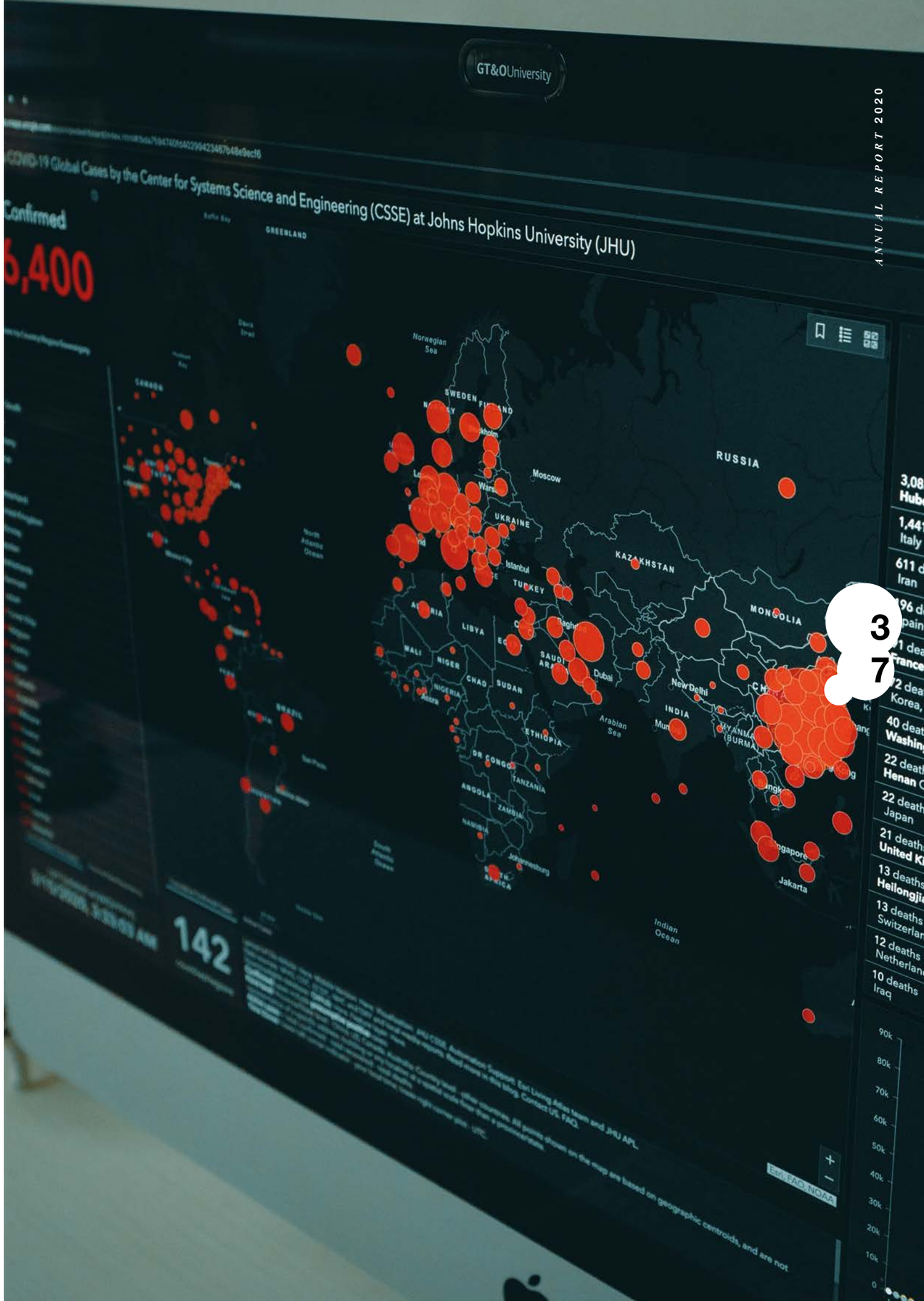
- Prof. Dr. Georg Groh, Department of Informatics, TUM
- Dr. Dirk Brand, Stellenbosch University

Researchers

- Edoardo Mosca, Department of Informatics, TUM
- Tobias Eder, Department of Informatics, TUM

Applied Partners

- McElory Hoffman, PhD, Pralexis
- Dr. Johan Van der Merwe, Pralexis



METHAD – Toward a MEdical ETHical ADvisor System for Ethical Decisions

The aim of this project is to develop an algorithm that can help make ethically relevant decisions in the clinic – a machine that roughly does what clinical ethics committees usually do. If successful, this will not only relieve the increasing burden on these committees, but also allow for quicker decision making in medically overwhelming situations such as the situation the world is currently witnessing - the COVID-19 pandemic. The project is a collaboration between the Institute of History and Ethics of Medicine and Department of Electrical and Computer Engineering at TUM.

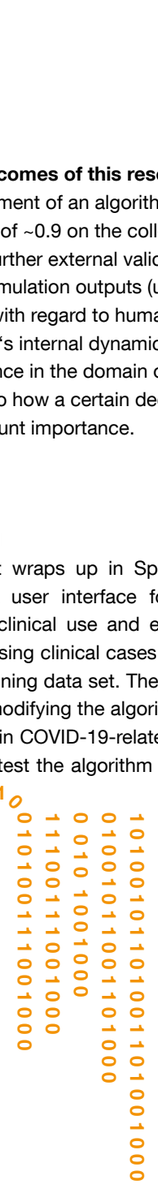
In 2020, the research team completed a framework to convert a wide range of ethics cases into machine-readable form, as well as a training data set of medical ethics textbook-cases and counter-factually augmented data. From this, they implemented pipelines for machine learning models and are currently working on developing a user interface.

The initial outcomes of this research include

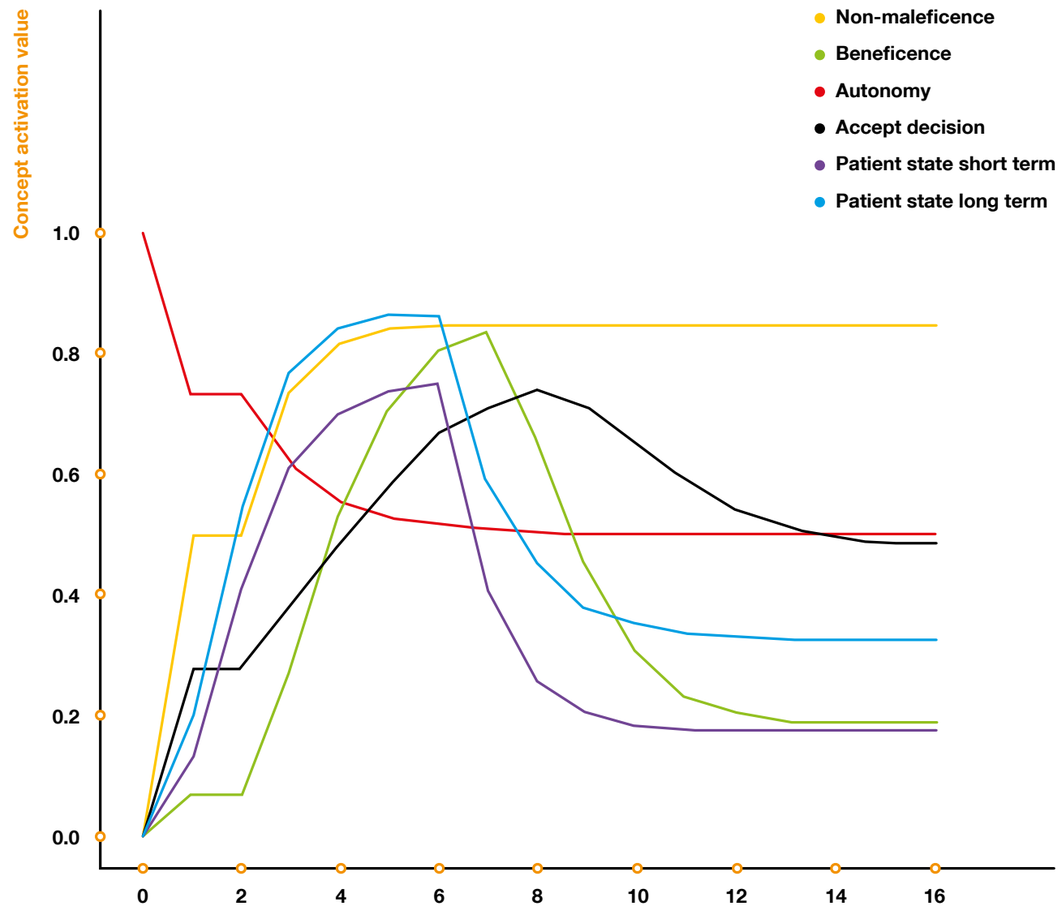
- The development of an algorithm that reaches an accuracy of ~0.9 on the collected data set (subject to further external validation going forward).
- Promising simulation outputs (using qualitative inspection) with regard to human comprehensibility of the model's internal dynamics. This is particularly important since in the domain of clinical ethics, transparency as to how a certain decision was reached is of paramount importance.

Plans for 2021

As the project wraps up in Spring 2021, the team will be creating a user interface for the algorithm that is conducive to clinical use and evaluating the algorithm's performance using clinical cases that were not included in the original training data set. They also plan to explore the possibility of modifying the algorithm in a way that permits its application in COVID-19-related decision-making in the clinic and will test the algorithm in clinical practice if time permits ... 0 0 1



Simulation results for case 4



2020 Papers and Projects Highlights

- Artificial Intelligence for Clinical Ethics Decisions: Ethical and Technological Challenges of Algorithmic Implementation (L. Meier, A. Hein, K. Diepold, A. Buyx – under review)
- Fuzzy Cognitive Maps with Flexible Activation Functions for Support in Decision Making in Medical Ethics (working paper)
- Artificial Intelligence for Clinical Ethics Committees (working paper)

2020 Conferences

- The Responsible AI Forum (TRAIF) Preview 2020, November 2020

Principal Investigators

- Institute of History and Ethics of Medicine, TUM
- Department of Electrical and Computer Engineering, TUM

Researchers

- Martin Gottwald, Department of Electrical and Computer Engineering, TUM
- Alice Hein, Department of Electrical and Computer Engineering, TUM
- Dr. Lukas Meier, Institute of History and Ethics of Medicine, TUM

2021 will bring us new challenges.

However, it will also bring us new opportunities.

Public Trust in AI and the Ethical Implications: A Comparative Study of Governmental Use of AI during the COVID-19 Pandemic

This project is a collaborative initiative of the Global AI Ethics Consortium (GAIEC), led by the IEAI. It uses a comparative case study methodology, examining the process and conditions surrounding specific countries' technology deployment choices during the COVID-19 pandemic.

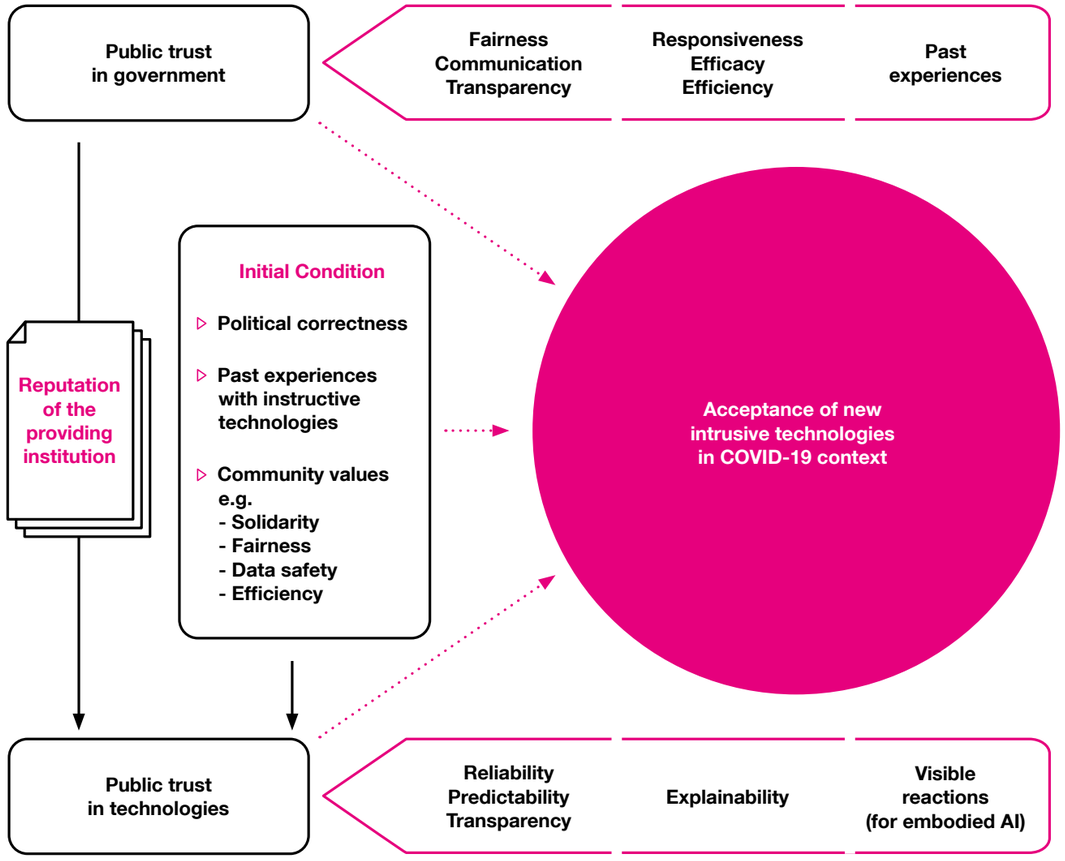
The present research expands upon the overview of this topic developed in the IEAI Research Brief: Ethics and the Use of AI-based Tracing Tools to Manage the COVID-19 Pandemic. By utilizing existing descriptive statistics, and survey data and conducting interviews of key experts in case study countries, the project aims to analyze and evaluate boundary alterations in terms of surveillance technology-related governance in times of public health crises and the corresponding government and public responses.

Finally, an estimation of the possible long-term effects and risks that those changes bring in terms of public acceptance, trust in governments and trust in technology will be assessed. The five countries studied include: China, Germany, Rwanda, Brazil, and Israel.

From the presented investigation, a reflection on the 'new normal' will be proposed. Moreover, policy and ethical recommendations and lessons learned will be presented for the attention of governments, policy makers and the private industry.

Plans for 2021

In 2021, the research team will collect data and interviews, analyze the data and organize a workshop on the topic at the Responsible AI Forum 2021. The goal of the research is to provide relevant information to practitioners, policymakers and developers. Thus, from the analysis and findings, the team plans to implement several outputs to reach a non-academic audience (science popularization videos, interactive websites broadcasting findings through interactive graphics, and policy / research briefs) to allow for easy and accessible overviews of the research outcomes. Moreover, publication will be prepared for academic journals. .



Principal Investigators

- Prof. Dr. Christoph Lütge, Institute for Ethics in Artificial Intelligence, TUM
- Caitlin Corrigan, PhD, Institute for Ethics in Artificial Intelligence, TUM

Researchers

- Auxane Boch, MSc, Institute for Ethics in Artificial Intelligence, TUM

2020 Papers and Projects

Ethics and the Use of AI-based Tracing Tools to Manage the COVID-19 Pandemic, IEAI Research Brief, June 2020

The Ethics and Practice of AI Localism at a Time of COVID-19 and Beyond

This new year-long project is part of the IEAI's efforts to promote research on the use of AI to manage pandemics, while building upon our partnerships with our Global AI Ethics Consortium (GAIEC) Members. The research is developed in a collaboration between the TUM School of Governance, The GovLab and the Alliance for Public Interest Technology at New York University and the Centre for Artificial Intelligence and Digital Ethics at the University of Melbourne.

As they grapple with COVID-19 and other challenges, cities around the world are mobilizing AI technologies to trace patient contacts, automate decision-making and monitor citizen movement and compliance with social distancing and quarantine. These developments are part of the emergence of AI Localism, a term coined by two of the project leads, Stefaan Verhulst and Mona Sloane for a new and radical shift of AI governance from the national to the local level. While local AI leadership offers greater agility and potential for innovation, the urgency of the crisis has left little time to address ethical and human rights questions and inform city officials of potential risks. It has also left little time for international comparison in which COVID-19-related AI Localism approaches work most effectively.

This project aims to fill these gaps. It will compare COVID-19-related AI Localism in three cities and on three continents – Melbourne, Australia, New York City, USA and Munich, Germany – to identify different categories of successful and ethical AI Localism, in the context of COVID-19. At the same time, the project will identify risks and challenges, including problematic AI applications, public concerns and conflicts over AI use between cities and state and non-state actors. The project goal is to determine characteristics of success and share successful approaches that can be emulated by cities worldwide.

In 2020, the first two of three project phases were initiated. First, a comparative review of COVID-19-related AI technologies and new governance aproaches was conducted. Then, public perceptions of ethical dimensions of AI localism were identified and examples collected of local COVID-19-related AI technologies and governance approaches aligned or misaligned with these perceptions.

To this end, relevant AI Localism cases were identified in all three countries and preparations were made for experimental and comparative studies.

Plans for 2021

In 2021, studies will be conducted across all three of the chosen sites. Different categories of successful and ethical AI Localism in the context of COVID-19 will be identified and ultimately a toolkit and training modules for city officials will be developed to help them advance local leadership in COVID-19-related technology and governance.

Principal Investigators

- PD Dr. Matthias Uhl, TUM School of Governance
- Prof. Stefaan Verhulst, The GovLab and the Alliance for Public Interest Technology, New York University
- Prof. Jeannie Marie Paterson, PhD, Centre for Artificial Intelligence and Digital Ethics, University of Melbourne
- Mona Sloane, PhD, The GovLab and the Alliance for Public Interest Technology, New York University

Researchers

- Andrew Young, The GovLab, New York University

Rule of Law, Legitimacy and Effective COVID-19 Control Tech

This collaborative project between TUM, the Centre for Artificial Intelligence and Data Governance (CAIDG) at Law School Singapore Management University and the British Institute of International and Comparative Law (BIICL) at the Bingham Center for Rule of Law is one of five new year-long projects for the IEAI aimed at promoting research on the use of AI to manage pandemics. It also aimes to build our partnerships within the Global AI Ethics Consortium (GAIEC). The project interrogates questions of rule of law protections in the course of the creation, implementation and maintenance of COVID-19 control technologies.

A number of nation state jurisdictions, in differing forms and at different trajectories, use a variety of technologies to control the spread of COVID-19. Many of these involve the compromise of personal data and the restrictions of individual liberties. These intrusions have prompted community reactions ranging from compliance to hostility and resistance. Popular discontent with these potentially intrusive controls (in Europe and Asia) has focused on challenges to personal data protection, questions about equal application of the measures, as well as restrictions on freedom of movement and association.

In response to this discourse, this collaborative project addresses possible regulatory approaches to mediate such concerns through the lens of the rule of law. While the project focuses on the COVID-19 context, its outcomes are of relevance for other crisis contexts and different technologies as well. Thus, the focus is COVID-19 specific, but with an interest in what learnings can be distilled from this crisis to be applied to future scenarios. The project uses standardisation, legal and discourse analysis, as well as empirical methods, such as the app walkthrough method, to provide answers to the research questions.

The research team kicked off the project in September 2020, conducting monthly roundtable discussions between the collaborators to provide input to one another. They have developed several working papers and organized an event related to the topics of rule of law and technology and community disquiet during the COVID-19 pandemic.

Plans for 2021

In early 2021, the team will host an Online Workshop Day. The initial aim of 2021 research is to focus on disaggregating the rule of law into key safeguards that facilitate and promote legitimacy and public trust. This will work as a foundation for other parts of the project. They will also work with an ethnographic app walkthrough with a media go-along approach to investigate how and which rule of law aspects are prevalent in the design and functionalities of COVID-19 tracing apps. Finally, they will conduct an analysis of governance and oversight over COVID-19 technologies in different nation's jurisdictions, also in relation to trust and community disquiet. These findings will be published in individual publications and in form of a white paper.

2020 Papers and Project Highlights

- Examining Community Disquiet in The Age of COVID-19 (A. Wee, M. Findlay).
- AI and Data Use: Surveillance Technology and Community Disquiet in the Age of COVID-19 (A. Wee, M. Findlay).
- Organised Event: "The rule of law in the technological age"

Principal Investigators

- Prof. Dr. Christian Djeffal, Munich Center for Technology in Society / TUM School of Governance
- Prof. Mark Findlay, LLD, Centre for Artificial Intelligence and Data Governance; Law School, Singapore Management University

Applied Partners

- Dr. Julinda Beqiraj, Maurice Wohl Senior Research Fellow in European Law, Bingham Centre for the Rule of Law, British Institute of International and Comparative Law

Researchers

- Dr. Antonia Horst, Munich Center for technology in Society, TUM
- Camila Hidalgo Avila, Technical University of Munich
- Alicia Wee, Centre for Artificial Intelligence and Data Governance; Law School, Singapore Management University
- Jane Loo, Centre for Artificial Intelligence and Data Governance; Law School, Singapore Management University
- Nyasha Winberg, British Institute of International and Comparative Law
- Rowan Stennett, British Institute of International and Comparative Law

How can we use AI to understand, predict and mitigate potentially harmful online trends and behavior?

AI and online behavior.

Online Firestorms
and Resentment Propagation on Social Media: Dynamics, Predictability and Mitigation

Using AI-based techniques, social media platforms allow millions of people to interact with one another, gather information and form opinions. Very recently, however, negative phenomena such as fake news, hate speech and online firestorms have shaken our beliefs and hopes about the positive power of social media to their very foundations.

With social media being an integral element of everybody's daily life, enhancing fairness, minimizing biases and reducing negative dynamics on various online platforms is of crucial importance. This three-year project addresses the mathematical modeling of the formation and dynamics of opinions in large groups of interacting people on social media. In a joint effort by the Professorship of Computational Social Sciences and Big Data and the Chair for Applied Numerical Analysis at TUM, the project investigates the driving factors leading to negative dynamics at the social media group level. The fundamental goal is to reveal the possible relationship between the simple "social forces" acting at an individual level and the potential emergence of a global behavior. The goal of the project is to offer approaches on how to detect, react to and possibly mitigate these negative dynamics in an early stage. The results of the study will provide insights of ethical relevance by discussing responsibility, delegation and control mechanisms in human-AI interacting systems.

In 2020, the team focused on collecting data from 21 different online firestorms. These data sets typically contain between 2,000 and 20,000 firestorm-related tweets and ten times the amount of reference tweets. The team performed an extensive analysis of linguistic features and cues, such as the usage of pronouns.

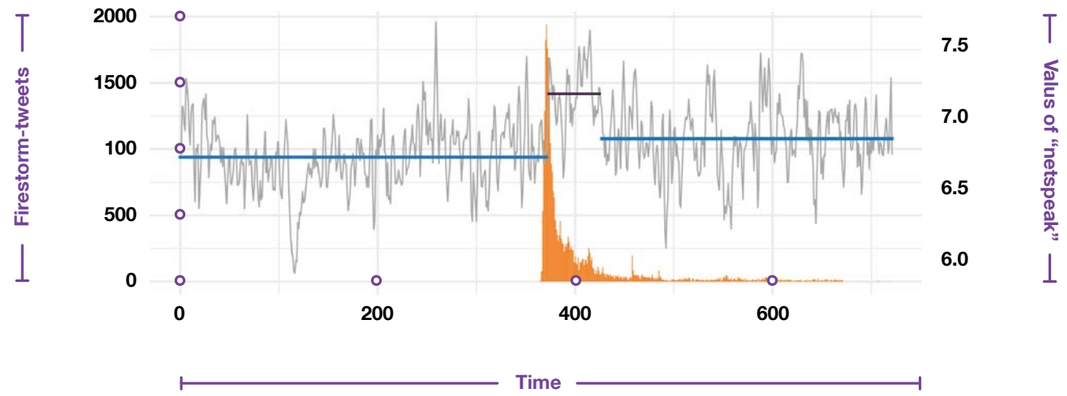
From this work, the team has found that

- linguistic features and cues have proven their usefulness in spotting outbreaks of online firestorms.
- this finding allowed for the development of a method to detect the point of change, systematically spotting on linguistic cues of tweets.

Building on these results, the team has extended the research into understanding the negative social dynamics of language change in social media networks. This allows for defining user-specific time series of both complete sets of words and well-motivated feature vectors, containing exclusively appearing pronouns or various other word classes accounting for linguistic cues. A next step is to visualize these dynamics. A further study undertaken in 2020 focuses on interacting multi-particle systems – consensus-based optimization. Based on this intuition from swarm intelligence, the team aims to find a global minimizer by considering a system of many interacting agents – subject to different forces – such as the urge to gather, improve the local position and explore the region.

Plans for 2021

In 2021, the team hopes to soon be able to join together our obtained knowledge tackling the obstacles of learning and inferring interaction rules and energies from data samples, with the performed data analysis of real-world Twitter data from multiple online firestorms. Furthermore, the project will extend data collection to other social media platforms.



Principal Investigators

- Prof. Dr. Massimo Fornasier, Department of Mathematics, TUM
- Prof. Dr. Jürgen Pfeffer, TUM School of Governance

Researchers

- Raji Ghawi, Computer Science, TUM
- Dr. Hui Huang, Department of Mathematics, TUM
- Konstantin Riedl, Department of Mathematics, TUM
- Dr. Mirco Schönfeld, Computer Science, TUM
- Wienke Strathern, Philology, TUM

2020 Papers, Projects and Achievements Highlights

- Learning Energies from the Evolutions of Critical Points (working paper) (C. Cipriani, M. Fornasier, H. Huang, K. Riedl)
- Against the Others! Detecting Moral Outrage in Social Media Networks, ASONAM (W. Strathern, M. Schönfeld, R. Ghawi, J. Pfeffer)
- A Convex Envelope Based Analysis of Consensus-Based Optimization (working paper) (M. Fornasier, T. Klock, K. Riedl)
- Understanding Social Dynamics of Language Change in Social Media Networks. (under review) (W. Strathern, M. Schönfeld, R. Ghawi, J. Pfeffer)
- Media coverage of topic in: ZDFzoom, ZDF, ZDFinfo and ORF2
- Participation in the TUM Data Innovation Lab, ca 26 projects (with ca 104 students) on several AI topics including Natural Language Processing
- Organization and advising of 17 Student projects (with 55 students) related to „Dynamics of Polarization and Radicalization" in two Computational Social Science project classes.

2020 Conferences

- Polarization on Reddit? Understanding Dynamics of User Interactions in Social Media Networks, INSNA Sunbelt 2020 – International Social Network Analysis Conference
- Against the Others! Detecting Moral Outrage in Social Media Networks, ASONAM 2020 – Conference on Advances in Social Network Analysis and Mining
- Plenary of M. Fornasier - Identification of Artificial Neural Networks, Istituto Nazionale di Alta Matematica (INdAM), December 2020
- TUM Exploratory Workshop on Geometric Deep Learning
- The Responsible AI Forum (TRAIF) Preview 2020, November 2020
- 12th annual Forum for Humane Economic Order "Menschenwürdige Wirtschaftsordnung" at the Academy for Political Education in Tutzing, March 2020
- Internal IT Cyber event of the European Central Bank

Online-Offline Spillovers – Potential real-world implications of online manipulation

People's online behavior often differs from their offline behavior. Research has shown that offline behavior may shape online personae. This project, a collaboration between the Department of Informatics and TUM School of Governance at TUM, analyzes the previously unexplored opposite question of whether online behavior spills over to the offline world.

Against the background of recent societal debates about fake news and online manipulations, this question deserves attention. In the project, the following research questions are addressed: How do online experiences affect offline behavior? And what mediates the respective effects? To this end, the research team employs experiments in the laboratory and in the field that address the question of whether, and how, online content influences people's beliefs. The intention is to be able to suggest measures that may help to mitigate tendencies for political polarization on social media.

Extremist and radical online content, for instance, can radicalize users' behavior, but it is unclear whether and under which circumstances it can also change users' attitudes and beliefs. In 2020, a study conducted as a Master's thesis measured how emotionally demanding reading radical posts can be, in order to assess whether such activity can influence users' attitudes on an affective level, using heart rate variability (HRV) as a measure of emotional load. A low HRV indicates persisting mental, cognitive or emotional stress. In the laboratory, a first set of 23 subjects were presented with radical posts (test condition) and neutral posts (control condition) on Facebook. During the whole process, the heart rate of participants was measured with wearable heart rate monitors.

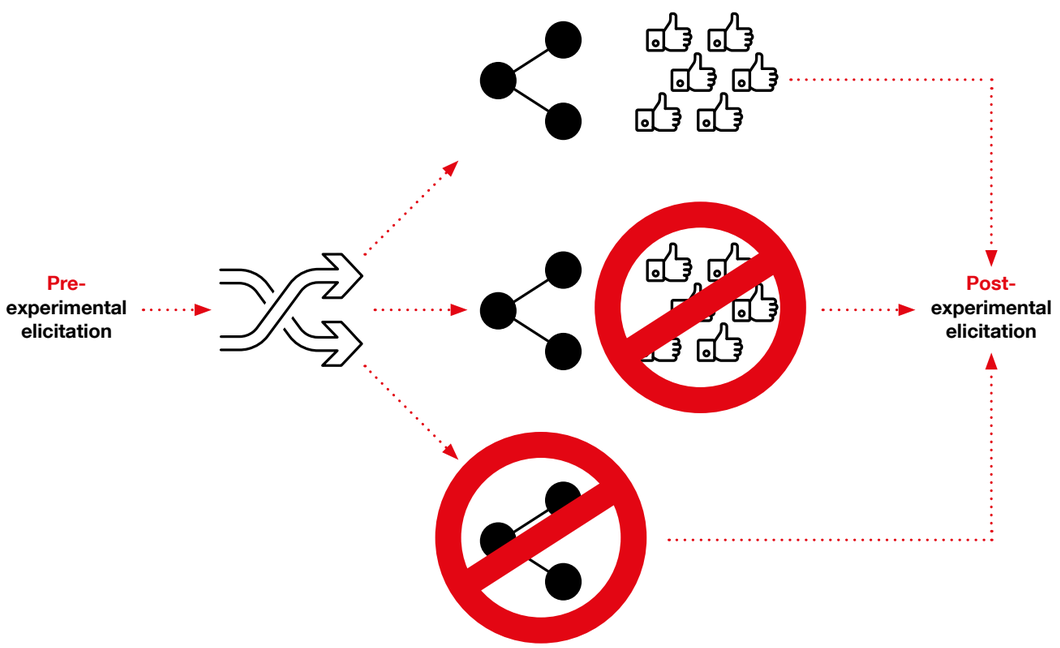
The team found that

- the initial data obtained tends to support the hypothesis that users would, on average, show a higher heart rate and lower heart rate variability (more mental, cognitive or emotional stress) when presented with radical posts.

However, more subjects are needed to obtain scientifically significant results. This is the task for 2021.

Plans for 2021

In 2021, the team will conduct a large-scale experiment on Twitter that will represent the core of the project. Participants will be randomly assigned to one of three treatments (see figure). In each treatment, subjects will be asked the same questions about their political attitudes at two points in time. The treatments that the subjects are exposed to vary, in terms of opinions they are asked to tweet about and feedback they receive. The expectation is that political polarization will be higher under circumstances where the subject is asked to interact more actively with the content.



2019/2020 Conferences

- Session on Ethics in Artificial Intelligence,12th Annual Forum for Humane Economic Order "Menschenwürdige Wirtschaftsordnung" at the Academy for Political Education in Tutzing, March 2020
- Polarization on Reddit? Understanding Dynamics of User Interactions in Social Media Networks, 40th Sunbelt Conference, July 2020
- Polarization on Social Media – An Experimental Approach, Workshop of Center Digitization, Technical University of Munich, March 2019

Principal Investigators

- Prof. Dr. Jürgen Pfeffer, Department of Informatics, TUM
- PD Dr. Matthias Uhl, TUM School of Governance

Researchers

- Dr. Gari Walkowitz, TUM School of Governance
- Wienke Strathern, Philology, TUM

Personalized AI-based Interventions Against Online Norm Violations: Behavioral Effects and Ethical Implications

Undesirable online behaviors such as hate speech or cyber-bullying have impacted public discourse, policy makers and platform providers. They can cause substantial harm to individual victims, limit the inclusivity of online settings and shift public discourse and perceived social norms in the online and offline world.

This multi-disciplinary research project combines the expertise from psychology, informatics and ethics (Max Planck Institute for Research on Collective Goods in Bonn and the Department of Informatics at TUM), to address the problem space of how AI-based interventions can be tailored (in personalized ways) and applied to counter online norm violations and to enhance their effectiveness.

The project investigates:

1. Whether interventions against online norm violations have differential effects on the attitudinal and behavioral reactions of transgressors, victims and witnesses of those violations depending on their personality and social identity.
2. The ethical ramifications from specific forms of AI-based personalized interventions for the norm transgressor and social media platforms in general.
3. The balance between the deployment of effective personalized interventions, the associated ethical considerations and the need for privacy and data protection, while also considering the applicability of existing technical methods for privacy-preserving AI to learn and deploy personalized interventions.

2020 activities and results of the project include:

- Using a sizable survey experiment, the team found that the perception of group norms predicted attitudinal and behavioral support for counter-speech against homophobic hate. These results suggest that opinion-based identities might become salient in the face of hate speech.
- Using a descriptive approach and a meta-ethical perspective, the team analyzed the impact of de-anonymization on a South-Korean social media platform, Naver. Naver's interventions in response to the South Korea's Network Act and platform-driven action against hate speech had a positive influence in reducing the amount of hate speech, although they also decreased the total number of comments written overall on the site. Our work suggests that ethical responsibility for online hate speech should not only lie in the hands of the government, but also on the platforms that host and enable hate speech between citizens.
- Using a comprehensive and systematic review of the technology landscape and literature, the team revealed that moderation on the internet and social media is predominantly reactive, mostly relying exclusively on the deletion of content and banning of users. This identifies a research gap concerning algorithmic moderation for custom communities (such as Facebook's Groups).

Plans for 2021

Based on these initial findings, ongoing work includes systematically iterating the study setup, and manipulating the content of counter-speech to, in particular, match or mismatch participants' political opinions regarding the issue. The aim is to gain a detailed understanding of how the content of counter-speech can match an online user's identity and develop guidelines to maximize interventions' effectiveness in terms of facilitating support for counter-speech and independent counter-speech to subsequent hate. Regarding meta-ethical challenges, the team is currently collaborating on a detailed discussion of transparency, censorship and redress, with the objective of providing forward-looking recommendations for the ethical design of future personalizing moderation tools. Finally, the team aims to develop guidelines on how to adapt algorithmic moderation technology to assist in context-specific personalized moderation. This work will be presented at the upcoming AAAI Spring Symposium on Implementing AI Ethics (February 2021), and an academic seminar at the University of Magdeburg (January 2021).

2020 Paper and Project Highlights

- How Theories of Personal Identity Reveal and Help Clarify Ethical Challenges in Social Media User Modeling (under review) (S. Engelmann, J. Grossklags)
- Ordinary People as Moral Heroes or Foes? Role Model Narratives in China's Social Credit System (under review) (M. Chen, S. Engelmann, J. Grossklags)
- A Review of Ethical Considerations for Moderation Tools on Social Media Platforms. (working paper) (T. Kuo, S. Engelmann, J. Grossklags)
- Naver Closing Pandora's Box: The Effect of Platform Governance on Cyber Harassment on Naver (working paper) (N.G. Kang, T. Kuo, J. Grossklags)

2020 Conferences

- Online Hate Speech and (Automated) Counter Speech, Digitale Woche, September 2020
- Personalized AI-based Interventions Against Online Norm Violations, The Responsible AI Forum (TRAIF) Preview 2020, November 2020
- Fighting Hate Speech: Group Norms and Intervention Behavior, Sociology and Psychology Seminar Series, Max Planck Institute for Collective Goods, October 2020
- Fighting Hate Speech: Group Norms and Intervention Behavior, Social Behavior and Decisions Lab Meeting, University of Virginia, November 2020
- AI, Risk Emergence & Zero Trust Networks, International Conference on Complex Systems, July 2020

Principal Investigators

- Prof. Dr. Anna Baumert, TUM School of Education and Max Planck Institute for Research on Collective Goods
- Prof. Jens Grossklags, PhD, Department of Informatics, TUM

Researchers

- Dr. Julia Sasse, Senior Research Fellow, Max-Planck-Institute for Research on Collective Goods
- Niklas Cypris, Max-Planck-Institute for Research on Collective Goods
- Severin Engelmann, Department of Informatics, TUM
- Felix Fischer, Department of Informatics, TUM
- Tina Kuo, Department of Informatics, TUM

Can everybody hear me?

AI, governance and regulation.

How can we manage trust and distrust in AI-based machine learning applications, where the inner workings are unclear and murky, even for experts?

TrustMLRegulation – Managing Trust and Distrust in Machine Learning with Meaningful Regulation

Meaningful regulation for technology is understood as consisting of regulatory frameworks, institutional designs and best practices of governance and oversight that are both politically and technically feasible.

To make these regulations effective, sufficient trust in the functional correctness of the regulated technologies is also needed, while at the same time still allowing for freedom to facilitate innovation.

To this end, this IEAI related project started out with the goal of exploring the border that divides the domains of trust and distrust in autonomous and intelligence systems. In particular, the collaboration between the Department of Electrical and Computer Engineering and the Munich Center for Technology in Society at TUM looked at the governance challenges for industrial corporations during the development of such systems, addressing the following questions:

1. What could be suggestions for effective regulation, certification and oversight of AI-based applications? Are they feasible in terms of how they fit into the existing institutional landscape? What are the limits of their application and are there ways to overcome such limits?

2. Are those suggestions for regulation and governance, that are institutionally feasible, also manageable on a technical level? How can selected machine learning methods be tested and evaluated in a way that fits the institutional and political requirements of such suggestions?

Potentially unintended consequences and several technical characteristics, such as unpredictability and lack of transparency, pose considerable challenges to the current governance infrastructure. This concern has been recognized and taken up by different regulatory initiatives ranging from industry-based self-governance guidelines, standard-setting bodies, academic communities and policy actors on national, international and global regulatory levels.

The team started by mapping these initiatives and identified common threads, as well as controversies and differences between them. This included AI standardization initiatives such as the IEEE AIS, ISO/IEC AIS, SAE AIS, TPC AIS, ITU AIS or the Industry for Open AIS, as well as the EU AIS (EC's European Group on Ethics in Science and NewTechnologies (EGE), Declaration of Cooperation on AI, Artificial Intelligence for Europe). From this mapping, the team undertook two policy analysis deep dives in a comparative manner.

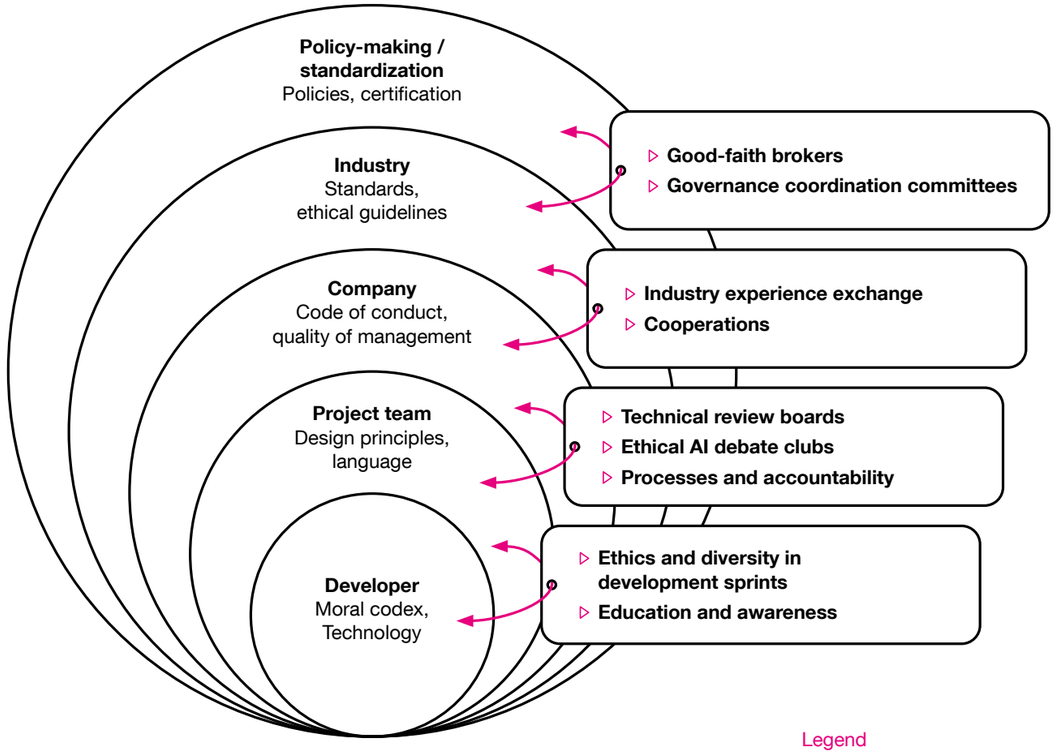
One analysis looked at industry self-regulation efforts in internal AI ethics guidelines in the insurance sector. The other focused on reconstructing the policy process that has led to the EU High-Level Expert Group on Artificial Intelligence recommendations and their implementations in EU Horizon funding and in tools such as the ALTAI framework.

Based on this work, the project had the following outcomes

- Proposals of a practical framework for multilevel governance of autonomous and intelligent systems (see figure). The framework identifies actors on five different levels of decision-making with a particular focus on industry actors, standardization and regulation, thus enabling the identification and development of potential tools for guidance on each level, as well as mechanisms for governance and oversight of the tools and technologies.

- Development of analogies to the governance of life sciences and agile approaches with good practices in these fields, exemplified using the case study of the discussion of the United Nations Group of Governmental Experts on Lethal Autonomous Weapons Systems between 2017 and 2019. Results show that linguistic analysis allows the tracking of changes, implying that a change of the topics, thus the focus of a governance process, can be achieved by the inclusion of actors from lower levels.

Feedback from four semi-structured interviews was used in the creation process of the governance framework. Interviews highlighted the benefits of the framework for identifying potential actors in decision-making processes and to frame governance processes, thus setting up effective governance structures for complex technologies such as autonomous and intelligent systems. Out of these findings, the team has submitted several proposals to fund further aspects related to this work, looking in-depth at effective levels of trustworthy and transparent AI.



Framework for multilevel governance of autonomous and intelligent systems

2020 Papers and Project Highlights

- A Practical Multilevel Governance Framework for Autonomous and Intelligent Systems (working paper) (L. Pöhler, K. Diepold, W. Wallach).
- A Sentiment, Trust and Content Analysis of the News Coverage on Artificial Intelligence in Germany (Master's Thesis, working paper) (J. Morandell)
- Governing Artificial Intelligence (Master's Thesis) (M. Pelepets)
- Governance of Autonomous and Intelligent Systems (AIS) – A Practical Multilevel Governance Framework (Master's Thesis) (L. Pöhler)
- AI Ethics in Insurance. Interpretation and Application of Ethical Principles (Master's Thesis) (C. Pai)
- Practices of Governing and Making Artificial Intelligence (Master's Thesis) (J.M. Samaniego)

2020 Conferences

- Digitization, Automation and Society, TUM-UQ online Workshop, July 2020
- Digital Superpowers and Geopolitics, TU Vienna Lecture Series, January 2021

Principal Investigators

- Department of Electrical and Computer Engineering, TUM
- Munich Center for Technology in Society, TUM

AI and sustainability.

What are the legal and ethical requirements for the use of big data and AI-based technology in Earth Observation?

Artificial Intelligence for Earth Observation (AI4EO): Reasoning, Uncertainties, Ethics and Beyond

In 2020, the IEAI became part of a €5 million BMBF grant for the newly established Future Lab Artificial intelligence for Earth Observation (AI4EO). Led by the Technical University of Munich in cooperation with the German Aerospace Center, the Lab brings 20 renowned international organizations across nine countries and 27 highly ranked scientists at all levels together to address three fundamental challenges in EO-specific cutting-edge artificial intelligence research – Reasoning, Uncertainties and Ethics.

The goal of the ethics segment of the AI4EO Future Lab is to support ethically mindful research and decision making by scientists at the very early stages of AI4EO research. Combining structured (ethics) research and unstructured/semi-structured interactions with other scientists working in the field of AI4EO, the aim is to create a practical approach by which scientists can (self) analyze their research through an ethical lens. The research includes studying whether existing (dominant) Western ethical theories and Eastern philosophical/ethical thought are adequate to devise such an approach, particularly in the light of uncertainties that dominate the rapidly expanding field of AI4EO.

Initial interactions and surveys conducted with AI4EO scientists revealed that:

- Although more than 40% of the scientists/researchers stated that they have read one or more major AI ethics guidelines, 90% found these guidelines to be not useful for their daily work and research.
- Yet, AI and ML models are becoming increasingly relevant in EO research as the quantity of data available for research increases, and more (including unconventional) sources of data are fused with EO/RS data.
- Many early-stage researchers are aware of well-known issues such as privacy, but are not aware of how to effectively and comprehensively manage privacy concerns. They are also unaware of any method or approach that can help them identify ethical concerns (and opportunities) at any stage of their research.
- Ethical concerns linked with AI4EO research remain largely unknown. Nonetheless, ethical issues are likely to come to light at much later stages of R&D. Accordingly, adoption of existing/known ethical theories (based either on duty, consequences or a combination thereof) and guidelines may not be adequate to help scientists identify, flag and address ethical issues in early stages of research, before it is too late.

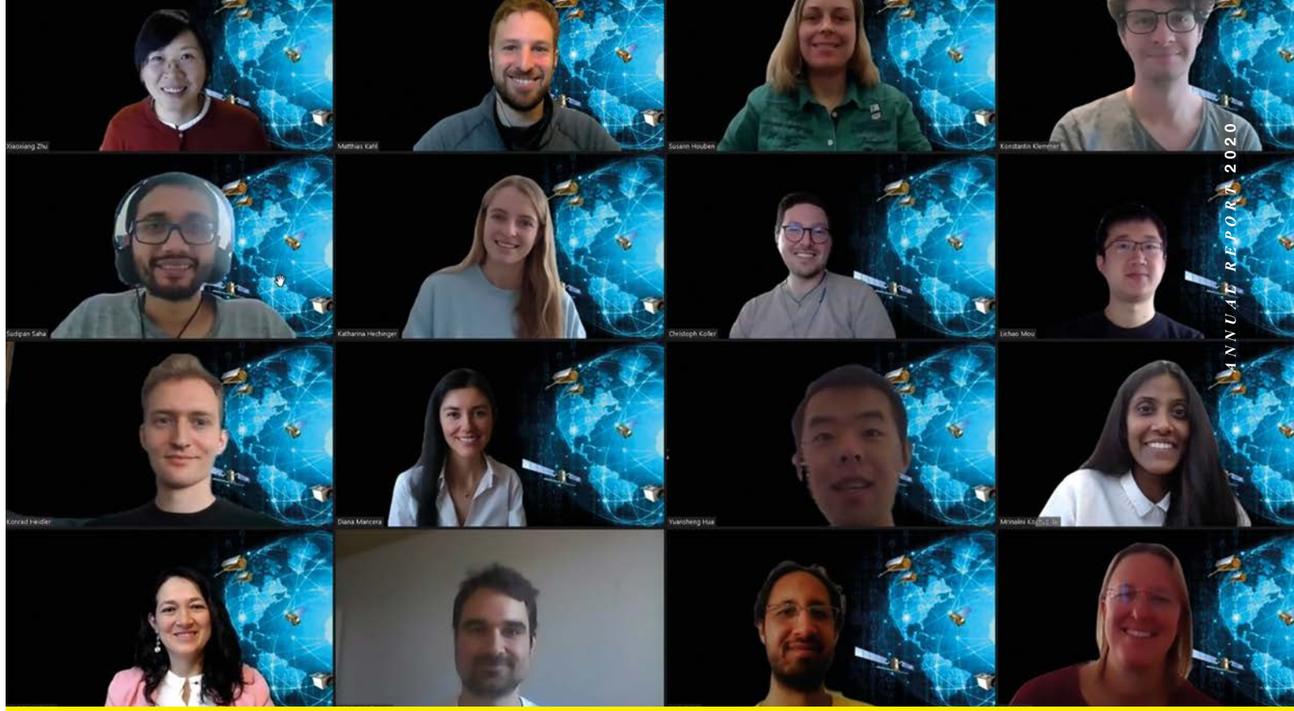
The research used preliminary inputs and feedback from various AI4EO scientists to create an initial outline of a novel approach and matrix for identifying ethical concerns that may arise or need attention at various stages of a typical AI4EO research and data lifecycle. Using a unique iterative methodology, this initial outline is now being continuously improved and its practical utility and applicability checked vis-à-vis various AI4EO research “cases”. This is done by using, testing and evaluating the matrix with inputs from AI4EO researchers/scientists at various research and career stages.

The research approach is adopted keeping the uncertain future and rapid technological progress in the AI4EO research sphere in mind. Therefore, the categories informing the matrix, as well as the ethical issues populating it, are not set in stone. Instead, an open-ended and introspective questionnaire guides the constant evolution and customization of the matrix for diverse research sub-fields and analytical methods. The ethics matrix for AI4EO research is not aimed at narrowing research innovativeness and freedom, but rather at supporting the development of ethics-oriented thinking amongst scientists already at early stages of their career.

Plans for 2021

In light of the fact that not all ethical issues can be addressed by technological solutions, it is necessary to develop training and educational curricula for scientists to support mindfulness and associated ethical decision making. For this, multi-disciplinary research engaging experts from human cognitive and brain sciences is planned going forward. Utilizing the “beyond fellows” program of AI4EO for this purpose, evidence-based techniques known to enhance mindfulness will be studied through appropriate experimental approaches to determine if they also support ethically mindful research and decision making, among scientists at various stages of their academic career.

The practical approach and training curriculum, it is hoped, will serve as “ethics orientation” for AI in Earth Observation in the future, particularly in the context of ensuring human dignity and supporting welfare-oriented human creativity and innovation. This includes freedom of choice and privacy on the one hand, and human values and responsibility to adequately ensure safety, as well as the continuation of beneficial innovation in the field of AI4EO, on the other.



2020 Paper Highlights

- Programming Away Human Rights and Responsibilities? “The Moral Machine Experiment” and the Need for a More “Humane” AV Future, NanoEthics (M. Kochupillai, C. Lütge, F. Poszler)
- Law, Business and Legitimacy. J.D. Rendtorff (Ed.), Handbook of Business Legitimacy: Responsibility, Ethics and Society, Springer-Nature (J. Brinkmann, M. Kochupillai)
- Incentivizing Research & Innovation with Agrobiodiversity Conserved In Situ: Possibilities and Limitations of a Blockchain-Based Solution, Journal of Cleaner Production (under review) (M. Kochupillai et al.)
- Outline of a Novel Approach and Matrix for Identifying Ethical Issues in Early Stages of AI4EO Research, IEEE IGARSS Symposium (under review) (M. Kochupillai)

2020 Conferences

- From Application-Agnostic to Application-Oriented AI4EO: Overview of Ethical Caveats and Opportunities, Keynote Talk at Munich Aerospace, July 2020
- Literature Review of Ethical Issues in AI4EO: Current Understanding and Scope for Improvement, AI4EO Lab Talk at Munich AI Future Lab, Technical University of Munich, November 2020
- Ethics and AI Enabled Decision Making: Towards More Affirmative-action and “Care” Oriented Data Labelling, The Responsible AI Forum (TRAIF) Preview 2020, November 2020
- Mindfulness, Mental Health and Ethical Decision Making in Times of Uncertainty, the Art of Living Germany, November 2020.

Principal Investigators

- Prof. Dr. Xiaoxiang Zhu, Department of Aerospace and Geodesy, TUM

Research Partners

- Prof. Dr. Richard Hans Georg Bamler, The Remote Sensing Technology Institute, DLR
- Prof. Dr. Massimo Fornasier, Department of Mathematics, TUM
- Dr. Mrinalini Kochupillai, Professorship for Signal Processing in Earth Observation, TUM
- Prof. Dr. Christoph Lütge, TUM School of Governance

Event Highlights

18
Public events

1,186
Attendees

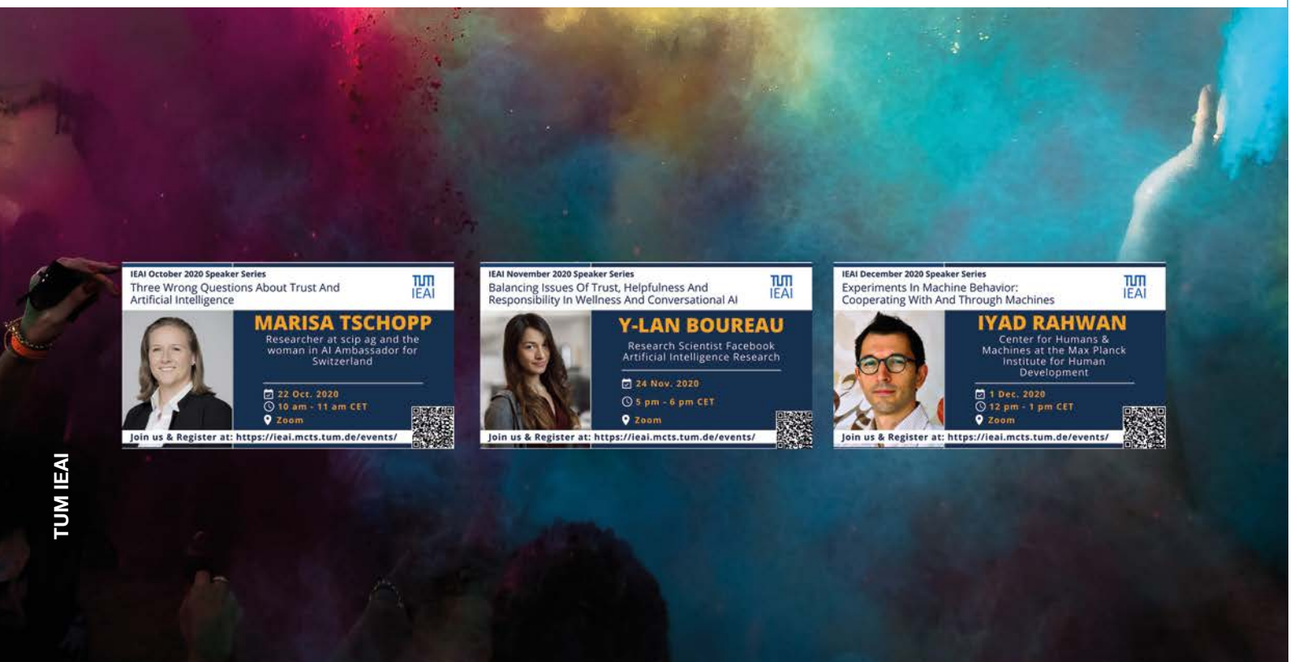
Speaker Series

With its Speaker Series, the IEAI invites experts from all over the world to talk about ethics and governance of AI. These events serve as an important platform for sharing new research and exchanging knowledge.

Since December 2019, the IEAI organized three in person and six virtual Speaker Series events, welcoming nine experts from three continents (Australia, Europe & North America). Our virtual events have allowed us to reach out to a broader audience from around the globe.
The talks ranged widely in terms of the topics and the speaker's backgrounds, with five of our speakers from academia, three from industry and one from a standard setting organization.

Some of the questions we had a chance to discuss with our distinguished guest speakers included:

- **Can we design fair AI for managing employees?**
- **How will AI change the way we think about mobility?**
- **Can AI ever be responsible?**
- **What kind of academic community do we need to build in order to meet the challenges of AI?**
- **Is it permissible for an AI tool to manipulate users in order to get them to do something that would be beneficial for them?**



Lionel P. Robert Jr.
The University of Michigan

★
Designing Fair Artificial Intelligence for Managing Employees

[December 2019]

Tae Wan Kim
Carnegie Mellon University

★
When is It Not Unethical for AI to Lie to Humans?

[December 2019]

Rolf Johansson
Autonomous Intelligence Driving GmbH

★
Autonomous Vehicles, Safety and AI

[February 2020]

Joanna Bryson
Hertie School

★
Artificial Intelligence is Nesessarily Irresponsible

[May 2020]

Clara Neppel
IEEE European, Vienna Office

★
Using Ethics Standardisation and Certification for Establishing Trust in the AI Ecosystem

[June 2020]

Huw Price
University of Cambridge

★
The Future of Artificial Intelligence: Academia's Role in Getting It Right

[July 2020]

Marisa Tschopp
scip AG and Women in AI

★
Three Wrong Questions About Trust and Artificial Intelligence

[October 2020]

Y-Lan Boureau
Facebook

★
Balancing Issues of Trust, Helpfulness and Responsibility in Wellness and Conversational AI

[November 2020]

Iyad Rahwan
Max Planck Institute of Human Development

★
Experiments in Machine Behavior: Cooperating With and Through Machines

[December 2020]

☆ ☆

TRAIF Preview 2020

Coming off celebrating our one-year anniversary, we had the opportunity to organize and host a major 2-day online event to showcase our work this year. The Responsible AI Forum (TRAIF) Preview 2020 was a sneak-peek into how the IEAI and its partners, the Global AI Ethics Consortium and the Responsible AI Network Africa, are promoting a sustainable, inclusive and comprehensive framework for the use of AI that delivers global benefit.

On the 12th and 13th of November, more than 20 experts from around the globe met virtually and shared their views, thoughts and research findings with an international audience. More specifically, Session 1 was devoted to the winners of the 2020 Call for Proposals for multi-disciplinary research on Ethics of the Use of AI to Manage Pandemics and Health-related Crises. During Session 2, researchers with projects under the umbrella of our AI and Online Behavior and AI, Choice and Autonomy research clusters gave insights into their initial work and findings. The last Session of the first day brought together five experts from Africa to discuss the opportunities and the ethical challenges for Responsible AI in Africa.

In April 2020, the IEAI joined forces with academic institutions, research centers and distinguished members of academia worldwide to launch the Global AI Ethics Consortium. During Session 4, several members of the Consortium had the chance to discuss the role of exchange and partnerships to promote a global perspective of AI ethics. The closing session of our TRAIF Preview provided a glimpse into how the IEAI is researching ethics and AI-enabled decision-making in practice. Five researchers illustrated how IEAI's projects engage with ethics and decision-making in key research areas, such as Mobility and Safety, Future of Work, Sustainability as well as Medicine and Health Care. This two-day virtual event was also a teaser of The Responsible AI Forum (TRAIF 2021) that will take place in December 2021 in Munich.

This international forum aims to bring together members of industry, civil society, government and academia to discuss the most relevant and pressing issues related to the responsible use of AI through shared stories, cutting-edge research and practical applications. TRAIF 2021 also aims to encourage exchange between research and practice through productive discussion and demonstration.

responsibleaiforum.com

Other Events

In addition to our Speaker Series and the Responsible AI Forum Preview, the IEAI hosted a workshop in January 2020 to discuss the new guidelines issued by the Data Ethics Commission in October 2019.

Paul Nemitz, Principal Adviser to the Directorate-General for Justice and Consumers of the European Commission in Brussels, spoke at length about his close involvement in drafting the new guidelines and what he sees as the most important challenges and considerations in governing AI today. Extensive discussion followed with academics, students and members of industry, media and civil society, focusing on questions such as: What is the answer to governing such a space? What is the role of ethics? What is the role of regulation and how can we keep up with the pace of technology, given the deliberative nature of democracy?

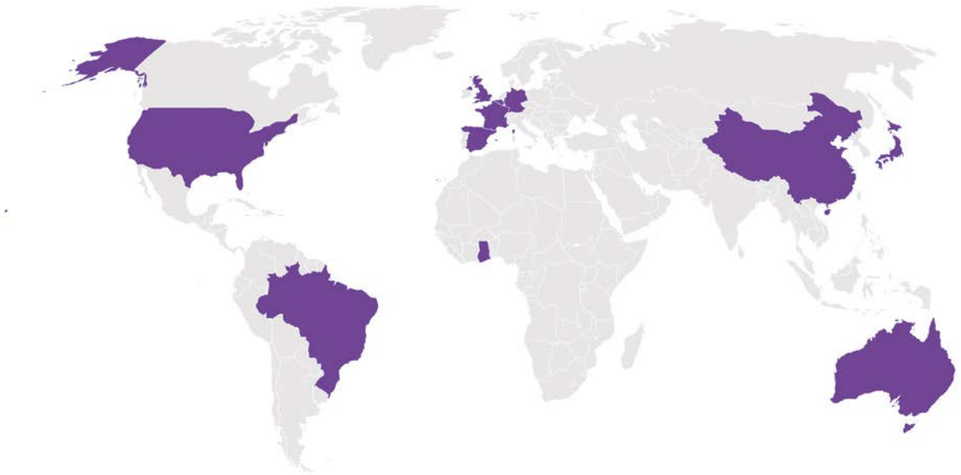
During its first year of operation, the IEAI co-hosted several events, such as four RAIN-Africa Workshops (discussed in our Networks section) and co-organized other forums, such as a panel workshop with the Brussels-based group FIRPA on "Data and Ethics in AI: How will German EU Presidency Shaper Prospects?" and the 12. Forum für Menschenwürdige Wirtschaftsordnung: Ethik in der Künstlichen Intelligenz in cooperation with the Akademie für Politische Bildung Tutzing. During this two-day event, renowned experts from the fields of philosophy, economics, law and technology discussed how the intelligence systems could be designed humanely and ethically.

Global AI Ethics Consortium

As TUM's strategic partnership initiative articulates: "No single university or country can master today's scientific challenges on its own." Working on a local, national or even regional basis on these matters is no longer enough. That is why the IEAI has led the way in the creation of the Global AI Ethics Consortium (GAIEC), joining forces with academic institutions, research centers and distinguished members of academia worldwide to foster trust in data and technology and maximize the potential of AI while limiting its harms.

The GAIEC started as a network of experts interested in examining the responsible use of AI in the fight against the COVID-19, but has since expanded to include other questions related to AI ethics. Its activities include shared research development and outreach, as well as co-promotion and participation in the upcoming "AI Ethics – Global Perspectives" online course (co-developed by the IEAI and the GovLab). Through this work, the GAIEC aims to bring a global awareness of the importance of ethics in AI, as well as acknowledging the diverse contexts around the globe in which AI is employed.

The GAIEC has continued to grow from its initial members to currently 20 experts from academic and research institutions on six continents.



The time
for analyzing
how AI is deployed
- whom it affects, how it affects them, what its broader social and economic impacts are -
is now.

[Excerpt from the GAIEC State of Purpose]

Additionally, in an effort to expand the scope of research on AI and its use in health-related crises, and to further promote partnership with the GAIEC, the IEAI supported four new multidisciplinary research projects starting in fall 2020 with a total of €400,000 in funding. In the call for proposals, the IEAI promoted submissions for projects that involved partners from the GAIEC and their institutions, with the vast majority of the winners having a GAIEC partner as a project Principle Investigator. The projects cover a range of topics related to the responsible use of AI to manage pandemics and health-related crises (see Research Project Reports for details).

In order to ensure the applied nature of the results of the projects, each project also has external partners in industry, at policy institutions or in government. In addition, the IEAI is directly carrying out a project that uses a comparative approach to examine the use of AI-enabled tools used around the world during the COVID-19 pandemic, specifically focused on combining the research efforts of GAIEC partners.

ETHAN
Ethical AI for Pandemic Management

TUM & Stellenbosch University

Rule of Law, Legitimacy and Effective COVID-19 Control Technologies

TUM & Singapore Management University

Public Trust in AI and the Ethical Implications: A Comparative Study of Governmental Use of AI During the COVID-19 Pandemic

TUM & GAIEC

Building Strategic Partnerships to Understand Ethics and the Use of AI to Manage Health-related Crises

TUM, Imperial College London & University of Tokyo

The Ethics and Practice of AI Localism at a Time of COVID-19 and Beyond

TUM, The GovLab at NYU & University of Melbourne

Responsible AI Network (RAIN)-Africa

The transnational nature of AI technology platforms makes the ethical application of AI a global concern. Because AI inherently interacts with its surroundings, cultural, political and environmental differences, in context it may produce different sets of ramifications (positive or negative) resulting from the use of AI-based technology. There is, therefore, a growing need to understand how AI may impact or be accepted by society in various regions around the world.

To this end, The Responsible AI Network – Africa (RAIN-Africa) was founded through a partnership between the Faculty of Electrical and Computer Engineering at Kwame Nkrumah University of Science and Technology (KNUST) in Ghana and the IEAI in 2020.

The aim is to build a network of scholars working on the responsible development and use of AI in Africa and to bring together emerging researchers to discuss and develop joint projects on the ethical and social challenges arising at the interface of technology and human values, specifically as it relates to sustainable development.

With reference to the fourth industrial revolution, the way to go is to really focus on AI and how we can responsibly use it for economic development within the Africa sub-region. To be able to do this, it is important to build capacity and bring together like-minded people that can work together in pushing the agenda forward. And, I think this is what the Responsible AI Network – Africa stands for: where we have like-minded intellectuals, like-minded people in AI, working together in various projects. It is important that we look at the ethical use of it within the sub-region, so that people are not disadvantaged.

[Prof. Jerry Kponyo]
Head of the Quality Assurance and Planning Unit of the Kwame Nkrumah University of Science and Technology (KNUST), Co-Founder of the RAIN-Africa and member of the GAIEC.

Through workshops and exchange, the Responsible AI Network – Africa provides a platform to promote cross-regional and interdisciplinary research collaboration.

rainafrica.org

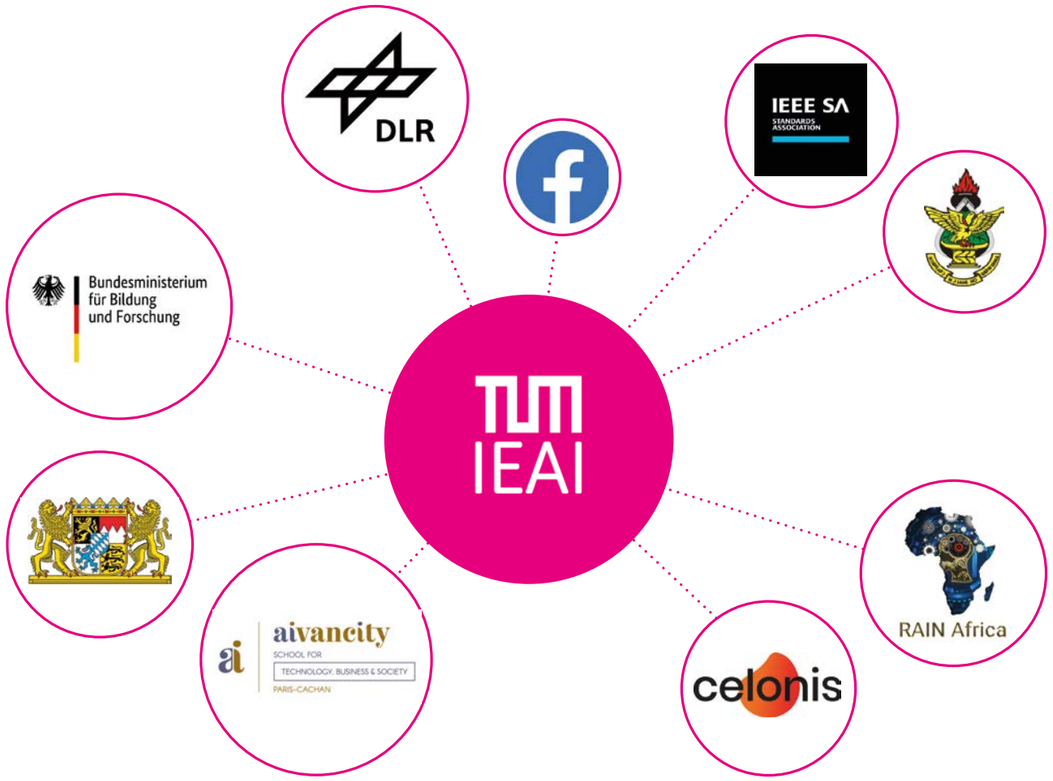
In 2020, RAIN-Africa held four online workshops, plus a kick-off event in Ghana, developed new partnerships with FAIR Forward (GIZ) and the Makerere AI Lab in Uganda, recruited ten country representatives from across Africa, launched of RAIN-Africa's website and developed its social media networking platforms. Our research assistant, Laud Ammah, has worked closely with our partners across Africa to strengthen these activities. In the first virtual workshop in May, the coordinators of the network, Dr. Catlin Corrigan, Prof. Jerry John Kponyo and Prof. Christoph Lütge, had the opportunity to stress the importance of this partnership to increase technological development and its effectiveness by making it more relevant to the African content. This event was followed by workshops focusing on AI ethics in different applications, including agriculture and healthcare, as well as the state of AI ethics in higher education in Africa (co-hosted with the GIZ FairForward initiative).

- **May 2020**
  Ethics and Artificial Intelligence
- **July 2020**
  Responsible AI and Agriculture
- **September 2020**
  Integrating AI Ethics into Higher Education Curricula in Africa
- **December 2020**
  Responsible AI and Healthcare

We ended the year by collaborating with Montreal AI Ethics Institute in December to co-host the event on Perspectives on the Future of Responsible AI in Africa, further establishing our growing network as we continue our work in 2021.

The IEAI uses both its networks and partnerships to enhance its mission of promoting research on AI ethics and providing a platform for meaningful cooperation between the wide range of important stakeholders in the AI ethics field. In order to leverage its research capabilities, attract international funding and get access to a larger pool of talented experts and students, the IEAI develops systematically strategic partnerships with other universities, academic centers, governmental bodies, civil society and industry.

In addition to working with partners from the Responsible AI Network – Africa and the Global AI Ethics Consortium, the IEAI is working with partners from the German Aerospace Center (DLR) through the Artificial Intelligence for Earth Observation (AI4EO) Future Lab. Awarded five million euros by the German Federal Ministry of Education and Research (BMBF), the Lab was one of only three to receive funding in Germany. We are expanding our work with applied partners as well. In January, the IEAI signed a Letter of Intent to work with the IEEE. In July, the IEAI signed a Letter of Intent with Celonis. These letters formalized the collective goal to promote AI ethics and bridge the gap between research and practice on responsible AI. The Celonis/IEAI partnership initially focuses on AI in healthcare through a joint Master's thesis project. In November, the IEAI also signed a Letter of Intent to work with the new school in France, aivancity, focusing on promoting student and learning exchanges going forward.



In 2021, we plan to continue our work on new partnerships in the pipeline and announce new partners soon. Through these partnerships, the IEAI aims to build a strong network of world-class scientific and corporate collaborators to address AI ethics-related challenges.

Publications and Outreach

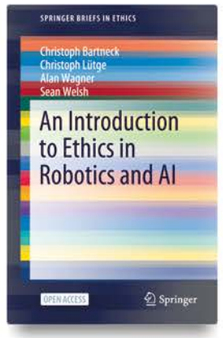
Outreach is an incredibly important component of the IEAI mission. Research on AI ethics needs to translate to practice. Otherwise, the impact of our work would remain minimal. Therefore Publications and outreach about the results of our research make up a core task of our work.

To this end, the IEAI itself produces a bi-monthly Research Brief Series. The aim of these publications, developed by our research staff, is to highlight important topics in AI ethics and explore the practical ethical considerations related to them. These succinct reviews are aimed at a broad audience with the goal of informing the public on the role of AI in society and the ethical challenges that accompany that expanding role. During 2020, the IEAI team published six Research Briefs.

In addition to our in-house publications, our research groups have produced several thesis projects and academic works that are currently under review or in the publication process (See also Research Project Sections).

Of note, IEAI Director Christoph Lütge jointly published An Introduction to Ethics in AI, which was launched as an Open Access book in August 2020.

Moreover, our researchers have participated in conferences and other events around the world on the subject of AI ethics (See also Research Project Sections).



2020 IEAI Research Briefs

- **February** ▸ Data Ethics Commissions Guidelines
- **April** ▸ Ethical Implications of the Use of AI to Manage the COVID-19 Outbreak
- **June** ▸ Ethics and the Use of AI-based Tracing Tools to Manage the COVID-19 Pandemic
- **August** ▸ The Potential for AI in Implementing the Green Deal and Ethical Implications
- **October** ▸ AI Ethics: Why Does It Matter?
- **December** ▸ Negative Dynamics on Social Media and their Ethical Challenges on AI

Publication Highlights

AI4People: Ethical Guidelines for the Automotive Sector – Fundamental Requirements and Practical Recommendations
▸ C. Lütge, F. Poszler, A. Acosta, D. Danks, G. Gottehrer, L. Mihet-Popa, A. Naseer // International Journal of Technoethics, 2021

There is Not Enough Business Ethics in the Ethics of Digitization
▸ C. Lütge // Ethical Business Leadership in Troubling Times, 2020

Programming Away Human Rights and Responsibilities? The 'Moral Machine Experiment' and the need for a more 'Humane' AV Future
▸ M. Kochupillai, C. Lütge, F. Poszler // NanoEthics, 2020

Künstliche Intelligenz und maschinelles Lernen in der intensivmedizinischen Forschung und klinischen Anwendung
▸ A. Peine, C. Lütge, F. Poszler, L. Celi, O. Schöffski, G. Marx, L. Martin // Anästhesiologie & Intensivmedizin, 2020

Artificial Intelligence and Human Rights: A Business Ethical Assessment
▸ A. Kriebitz, C. Lütge // Business and Human Rights Journal, 2020

Against the Others! Detecting Moral Outrage in Social Media Networks
▸ W. Strathern, M. Schönfeld, R. Ghawi, J. Pfeffer // ASONAM, 2020

Artificial Intelligence for Clinical Ethics Decisions: Ethical and Technological Challenges of Algorithmic Implementation
▸ L. Meier, A. Hein, K. Diepold, A. Buyx // (under review)

How Theories of Personal Identity Reveal and Help Clarify Ethical Challenges in Social Media User Modeling
▸ S. Engelmann, J. Grossklags // (under review)

Outline of a Novel Approach and Matrix for Identifying Ethical Issues in Early Stages of AI4EO Research
▸ M. Kochupillai // IEEE IGARSS Symposium // (under review)

Understanding Social Dynamics of Language Change in Social Media Networks
▸ W. Strathern, M. Schönfeld, R. Ghawi, J. Pfeffer // (under review)

Conference highlights

JAN 2020 — Inaugural meeting of the ITU-T Focus Group on AI for Autonomous & Assisted Driving, FG-AI4AD

JUL 2020 — AI, Risk Emergence & Zero Trust Networks Panel, International Conference on Complex Systems

JUL 2020 — From Application-Agnostic to Application-Oriented AI4EO: Overview of Ethical Caveats and Opportunities, Keynote Talk at Munich Aerospace

SEP 2020 — Online Hate Speech and (Automated) Counter Speech, Digital Woche

MAR 2020 — 12th Annual Forum for Humane Economic Order "Menschenwürdige Wirtschaftsordnung" at the Academy for Political Education in Tutzing

JUL 2020 — Polarization on Reddit? Understanding Dynamics of User Interactions in Social Media Networks, 40th Sunbelt Conference

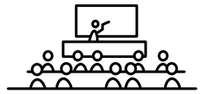
SEP 2020 — 3rd FAccTRec Workshop on Responsible Recommendation at RecSys

DEC 2020 — Indentification of Artificial Neural Networks, Istituto Nazionale di Alta Matematica

Teaching, Student Engagement and Resources

A key component of promoting AI ethics in a sustainable way is bringing the topic into the educational process. Through gathering information regarding courses offered at TUM that touch on AI and AI ethics and Master's or Bachelor's thesis opportunities related to AI, the IEAI further promotes the importance of AI ethics research and teaching by providing a useful platform for students and individuals interested in this field to get more involved. Moreover, our events are open to and well attended by students around the world.

Other Opportunities



Courses, lectures and seminars

Master's and Bachelor's Thesis

Which TUM Chairs and professorships are currently looking for students for topics on AI and Responsible AI? The IEAI offers an up-to date curated directory of multi-disciplinary AI ethics-related theses topics. In our efforts to bridge the gap between academia and practices, the IEAI also directly offers theses opportunities in collaboration with major companies and our AI ethics-related projects.

AI Courses, Lectures and Seminars

Before the beginning of each semester, students can find an overview of all of the AI-related courses, lectures and seminars offered at the TUM for the upcoming semester.

AI Ethics – Global Perspectives: Free Course, Co-developed by the IEAI and the GovLab

The Global AI Ethics Consortium (GAIEC), an IEAI led initiative, also supports the immediate needs for expertise and resources on the emerging ethical concerns, dilemmas and questions that arise as the use of AI grows. This support includes raising awareness and providing education regarding the ethical implications of decisions made on AI use. To this end, the IEAI together with its GAIEC partner, The GovLab, the Center for Responsible AI at NYU and the NYU Tandon School of Engineering have designed the course AI Ethics – A Global Perspective, a collection of short, free of charge online lectures and seminars from leading global experts in the field of ethical considerations for AI creation and use. This course will launch in early 2021.

aiethicscourse.org

Our aim is
to provide opportunities for students
to study, research, understand
and expand their skills
in the AI ethics field.

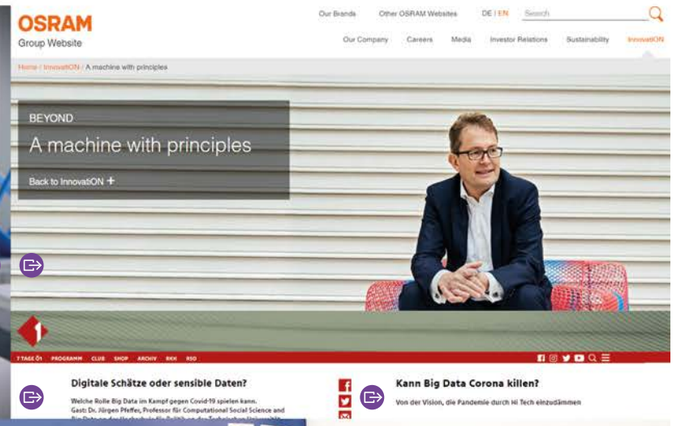
Master's and Bachelor's thesis

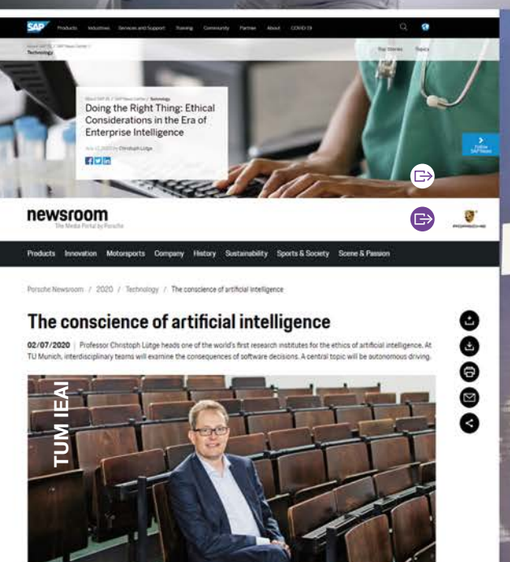
AI has the potential to impact our lives and our world like no other technology has done before. This potential comes along with a number of expectations around AI, important questions concerning its economic, societal and legal effects and an emerging discourse on AI ethics.

The public debate on AI ethics is fairly new and how we talk about artificial intelligence matters. The same applies to AI ethics. News media portrayals of ethical issues in AI play an important role in public understanding and perception. The IEAI seeks to take part in this debate and contribute to the broader conversations and concerns surrounding ethics and AI on an international level through our presence on social media, in the press and by producing content for public distribution. Through participating in interviews for magazine and newspaper articles, radio, TV and podcasts segments, the IEAI leadership, our researchers and the project Principle Investigators seize every opportunity to address the importance of AI ethics and its related challenges and prevent the spread of misinformation, making sure that ethics in AI get the attention it deserves.

Press

The IEAI, its research projects and its affiliated Principle Investigators have been featured across a multitude of press platforms since our launch. This speaks clearly to the quality and importance of our work and approach, and the growing interest in AI ethics outside academia.





[Osram ON]

The Innovation magazine: A machine with principles

Ethics is too important to leave to the responsibility of individuals.
And here, ethics does not primarily mean personal ethos –
as in the "respectable businessman" or the "responsible engineer."
It's about ethics at a system level, be it in companies or societies.
Some ethicists see it as mainly the legislator's duty.
But that's just one of many paths,
and it's long and hard.

[Porsche Magazine]

The conscience of artificial intelligence

We want to bring together all the important players
to jointly develop ethical guidelines for specific AI applications.
The prerequisite for this is for representatives
from the worlds of business, politics, and civil society
to engage in dialog with each other.

[SAP Magazine]

Doing the Right Thing: Ethical Considerations in the Era of Enterprise Intelligence

When we speak of ethics,
most people think about the moral principles
that govern the behavior of individuals or organizations.
Ethics can seem highly abstract, a subject more useful to ancient philosophers
than modern business and technology leaders.
Yet ethics is an important issue for today's decision-makers –
one that must be handled with extreme care.
The rampant growth of artificial intelligence and other intelligent technologies
in businesses makes ethics more than a philosophical question.
Ethics is now a form of risk management
that savvy businesses cannot ignore.

[Gesundhyte Magazine]

What the COVID-19 crisis has made clear is the need
for a coordinated, dedicated data and technology infrastructure and ecosystem for
tackling dynamic societal and environmental threats.
The crisis has also underlined the urgent need to discuss
the ethical considerations in the use of AI
in healthcare and public health contexts
and develop operational ethical frameworks
in the field of AI.

external link // only in digital version

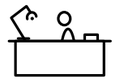
Q&A Series: Reflections on AI

In 2020, the IEAI also instituted the Q&A Series: Reflections on AI. This series was meant to engage with our invited speakers and IEAI affiliates in a more intimate, but also accessible way. During these monthly installments, we had the pleasure of speaking to eight experts and posing a set of basic yet important AI ethics-related questions.

In addition to the standard set of questions, each interviewee was also asked about their work. The Q&A Series attempts to further understand the importance of AI ethics, the dilemmas and misconceptions associated with AI and the role of the various stakeholders in the ethics and governance of AI, as viewed by the experts themselves.

What is the biggest misconception about AI?

What is the most important question in AI ethics right now?



Who should be in charge or involved in developing ethical frameworks and standards for AI?

What is the role of academia, research institutions and other centers when it comes to the ethics and governance of AI?

We often say that AI is changing or transforming the world. To what extent is AI changing us as humans?

I don't think we are ever going to have anonymity again.

▸ **Joanna Bryson**
Professor of Ethics and Technology, Hertie School

AI cannot solve humanity's old problems. It is humans together with AI that can solve them.

▸ **Clara Neppel**
Senior Director, IEEE Vienna

The effort of making sure technologies are beneficial, needs to be a globally collaborative one.

▸ **Huw Price**
Bertrand Russell Professor of Philosophy & Fellow of Trinity College University of Cambridge

There will always be the need for human intervention.

▸ **Jerry Kponyo**
Dean of Faculty of Electrical and Computer Engineering, KNUST

AI is a global challenge, so is AI ethics.

▸ **Christoph Lütge**
Professor of Business Ethics at TUM & IEAI Director

One of the most critical questions is the controversial and very critical relationship between the topic of trust and AI.

▸ **Marisa Tschopp**
Researcher at scip AG & Women in AI Ambassador for Switzerland

The biggest promise of AI in wellness is scalability.

▸ **Y-Lan Boureau**
Research Scientist at Facebook AI Research

We hope that our free online course will provide insights on what other countries have learnt from the last few years of experimenting with AI ethics and with AI governance.

▸ **Stefaan Verhulst**
Co-Founder & Chief R&D Officer of the GovLab at NYU

Social Media

More than
3,000
social media followers

Over the last years, social media has become an important medium of communication and social interaction. From the outset, the IEAI set a clear goal: to use these channels to build a strong and continuously growing community.

Through the systematic use of social media platforms such as LinkedIn, Twitter and YouTube, the IEAI team is able to present its work to a global audience. In these media:

- **we have been able to illustrate the progress of our multidisciplinary projects;**
- **communicate consistently with our community;**
- **give and receive valuable feedback;**
- **raise awareness regarding AI ethics-related issues;**
- **grow interest in the field of AI ethics;**
- **share interesting news and articles;**
- **promote the work of other members of our network;**
- **open new opportunities for collaboration with other organizations and**
- **generate a global conversation on the importance of AI ethics.**

The use of these means of communication helps the IEAI not only to promote its work to the outside world, but also to build an AI ethics culture both internally and externally. During its first year of operation, the IEAI devoted time and energy to creating strong bonds with its virtual community members and managed to attract a high level of attention on social media. This interactive relationship involves more than 1000 Twitter and more than 2000 LinkedIn engaged followers.

Our goal in 2021 is to keep growing our community and engaging more people in the AI ethics discussion through these platforms.

LinkedIn 2020

275
Posts

1,690
New followers

7.23%
Engagement rate

135,475
Impressions

Some of the most popular posts

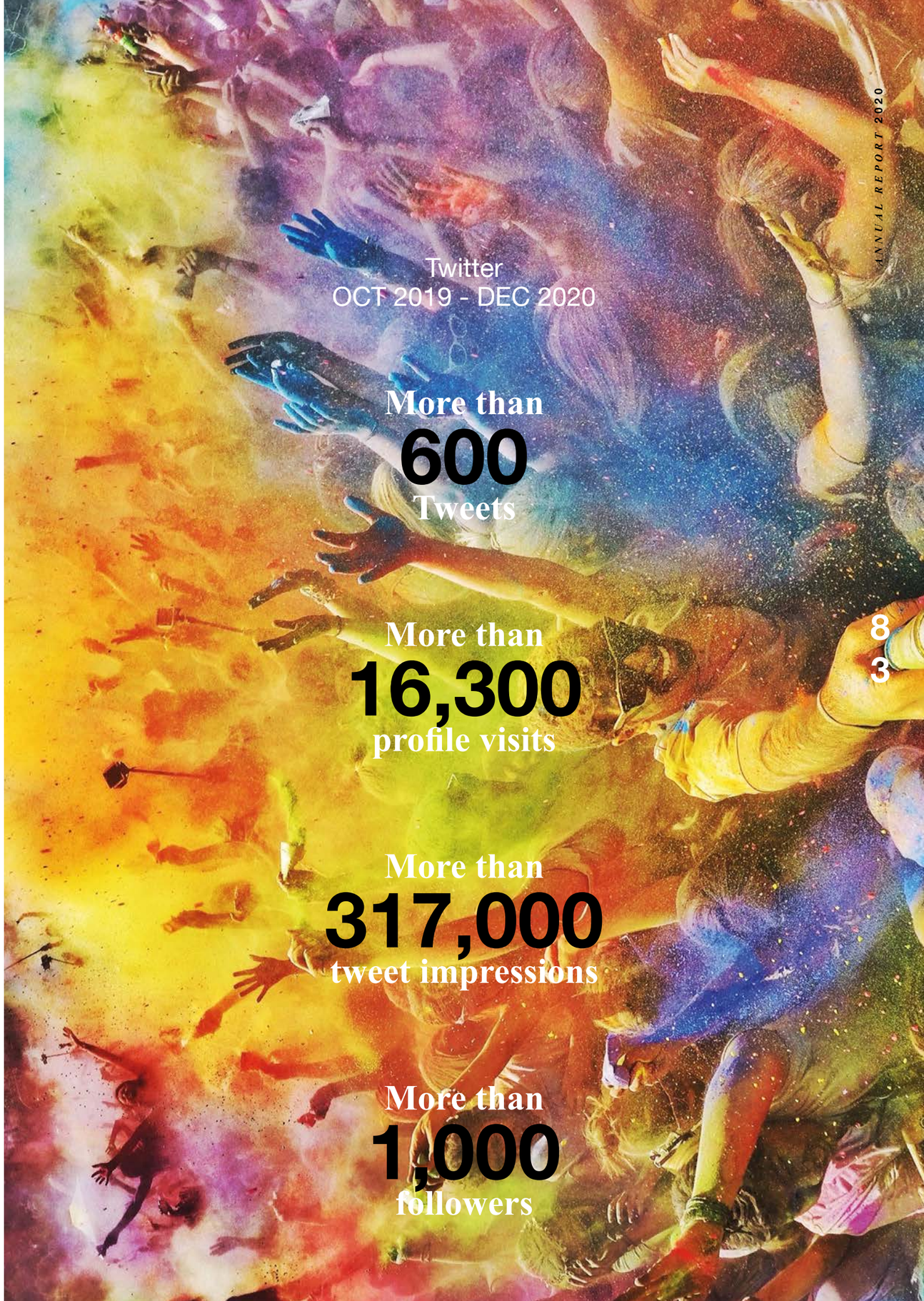
An Introduction to Ethics in Robotics and AI, open access book

October 2020 IEAI – Research Brief: AI Ethics: Why Does It Matter?

Reflections on AI, September 2020 – Q&A with Prof. Dr. Christoph Lütge

June 2020 IEAI – Research Brief: Ethics and the Use of AI-based Tracing Tools to Manage the COVID-19 Pandemic

August 2020 IEAI – Research Brief: The Potential for AI in Implementing the Green Deal and Ethical Implications



ieai.mcts.tum.de

TUM

Institute for Ethics in Artificial Intelligence

ANNUAL REPORT

2020

Contact Information
The IEAI annual report is published in English.
More information about the IEAI can be found on the Web at ieai.mcts.tum.de.

Overall Responsibility
Prof. Dr. Christoph Lütge

Published by
Technical University
of Munich,
Institute for Ethics
in Artificial Intelligence

Visitor's address
Marsstrasse 40
D-80335 Munich

Postal address
Arcisstrasse 21
D-80333 Munich

Tel. +49 89 289 22901
Email: ieai@mcts.tum.de

Concept, design and production
Andreas Rintzner
www.RINTZNER.de

Photography
© Andreas Heddergott
© Sonja Herpich
© Unsplash
© Envato Elements

Technical University of Munich (TUM)
Institute for Ethics in Artificial Intelligence
Arcisstraße 21
80333 München

ieai.mcts.tum.de

IEAI's priority is the generation of global, egalitarian and interdisciplinary guidelines for the ethical development and implementation of AI throughout society.