Explicit Normative Reasoning and Machine Ethics
—ENoRME—

BMBF Research Proposal: International Future Labs ("Internationale Zukunftslabore")

Prof. Dr. habil. Christoph Benzmüller, Freie Universität Berlin

July 22, 2019

Contents

Deutsche Zusammenfassung

BMBF-Ausschreibung: https://www.bmbf.de/foerderungen/bekanntmachung-2377.html

Projekttitel: Explicit Normative Reasoning and Machine Ethics (ENoRME)

Antragsteller und Zukunftslabor Leitung: Prof. Dr. habil. Christoph Benzmüller

Beantragte Laufzeit: 36 Monate, 1/2020 – 12/2022

Angestrebter Standort und koordinierende Einrichtung: Berlin-Dahlem, FU Berlin, FB Mathematik und Informatik. Berlin – exzellenter Forschungsstandort, mit einer hochdynamischen Start-Up Szene, und nationales Zentrum von Politik, Medien und Kultur – ist ein perfekter Standort für das ENoRME-Zukunftslabor. Berlin ist zudem sehr attraktiv für Gastwissenschaftler und zur Ausrichtung von Tagungen.

Forschungs- und Entwicklungsvorhaben: Theorie, Entwurf, Implementierung einer "cloud-basierten on-demand" Plattform und Infrastruktur zur Modellierung und Automatisierung ethischer und rechtlicher Theorien; Methoden- und Technologieentwicklung im Bereich der Maschinen-Ethik; Experimente und Anwendungsstudien in ausgewählten Bereichen, z.b. Autonomes Fahren, Soziale/Kognitive Robotik, Banken- und Finanzsektor; weitere anwendungsorientierte Forschung an der Schnittstelle zwischen Maschinellem Lernen, expliziter Wissensrepräsentation und Deduktion; vielfältige Wissenstransfermaßnahmen, inklusive Lehrveranstaltungen, Workshops, Konferenzen, Tutorien und e-Learning (MOOC); Öffentlichkeitsarbeit und Community Management; enge Kooperation mit nationalen und internationalen KI-Initiativen.

Verwertungs- und Anwendungsperspektiven: ENoRME wird aktuelle Forschungsarbeiten zum Normativen Schließen, Automatischen Theorembeweisen und der expliziten, formal-logischen Repräsentation von ethischem und regulatorischem Wissen zusammenführen und verfügbar machen in einer leicht zugänglichen, cloud-basierten Forschungs-/Entwicklungs-/Anwendungsplattform; diese Plattform, in Kombination mit entsprechendem Lehr- und Lernmaterial, soll den Einsatz deduktiver Techniken zur rechtlichen und ethischen Kontrolle von KI-Systemen in der Praxis nachhaltig fördern; ENoRME ist konzipiert als direktes Bindeglied zwischen Forschung und Anwendung, und soll gleichzeitig den Grundstein legen für den nachhaltigen Aufbau eines internationalen Netzwerks zum Thema.

Mehrwert zu laufenden Forschungs- und Entwicklungsarbeiten: "Trustworthy AI made in Europe" — Motivation und Ausgangspunkt sind aktuelle, international führende Forschungsresultate der beteiligten Wissenschaftler/-innen in den Gebieten (i) Automatisches Theorembeweisen und Deduktion, (ii) Universelles Logisches Schließen, (iii) Modellierung ethischer und rechtlicher Theorien und Normatives Schließen, (iv) Maschinen-Ethik, (v) Robotik und (vi) Maschinelles Lernen. Diese Themen werden derzeit von den partizipierenden Wissenschaftler/innen isoliert oder in bilateralen Kooperationen in zumeist akademischen Kontext bearbeitet. ENoRME wird diese thematisch verzahnten Einzelaktivitäten zusammenführen und bündeln. Innovative, praktische Entwicklungen und Anwendungen sollen dadurch ermöglicht werden, die derzeit noch nicht realisierbar sind (aber grundsätzlich machbar).

Partizipierende Partnereinrichtungen und Wissenschaftler/innen: [a]

FU Berlin: (1) Prof. Christoph Benzmüller, (2) Prof. Raúl Rojas, (3) Prof. Daniel Göhring; **U Koblenz-Landau:** (4) Prof. Claudia Schon (Prof. Uli Furbach); **U Luxembourg (EU):** (5) Dr. Xavier Parent, (6) Dr. Alexander Steen (Prof. Leon van der Torre); **U Bergen (EU):** (7) Prof. Marija Slavkovik; **U Miami (US):** (8) Prof. Geoff Sutcliffe; **U Liverpool (UK):** (9) Dr. Louise Dennis (Prof. Michael Fisher); **U Campinas (BR)**: (10) Prof. Walter Carnielli, (11) Prof. Juliana Bueno-Soler; **U Buenos Aires (AR):** (12) Prof. Maria Vanina Martinez

Weitere akademische Partner Prof. Toby Walsh (UNSW Sydney, AU), Prof. Beishui Liao (Zheijang U, CN), Dr. Adam Pease (Infosys, Palo Alto, US), Prof. Jan Broersen (U Utrecht, NL), Prof. Stephan Schulz (DHBW Stuttgart), Prof. Ralf Romeike (FU Berlin)

Industrie & Forschungseinrichtungen: Latentine GmbH (KMU, DE), wizAI solutions GmbH (KMU, DE), LuxAI S.A. (KMU, LU), KPMG Lighthouse Luxembourg (LU), DLR (DE);

Zusammenarbeit in Berlin: Berlin Mathematics Research Center MATH+; JFK-Institute (FU Berlin)

Besondere Eignung der Wissenschaftler/innen für ENoRME: (1) Universelles Logisches Schließen, Deduktionssysteme, Formale Methoden, Maschinen-Ethik; (2) KI, Robotik, Maschinelles Lernen; (3) Autonomes Fahren, Robotik, Maschinelles Lernen; (4) Beschreibungslogiken, Automatisches Theorembeweisen, Maschinen-Ethik, Ontologien; (5) Normatives Schließen, Deontische Logiken, Ethische und Rechtliche Theorien; (6) Automatisches Theorembeweisen, Systementwicklung; (7) Kollektives Schließen, Maschinen-Ethik; (8) Infrastukturen und Evaluations- und Testumgebungen für Deduktionssysteme; (9) Autonome Systeme, Verifizierbare Systeme, Maschinen-Ethik; (10) Nichtklassische Logiken, Logikkombinationen, Parakonsistentes Schließen; (11) Probalistisches Parakonsistentes Schließen; (12) Schließen bei Unsicherheit & Inkonsistenz, Semantic Web.

Projektziele und erwartete Resultate • Cloud-basierte Werkbank und Infrastruktur für universelles Schließen; • Standardisierung und Benchmarking im Normativen Schließen; • Theorie und Implementierung neuer, relevanter Logikkombinationen; • Beispiele mechanisierter und automatisierter ethischer und rechtlicher Theorien; • Spezielle Technologien (Theorembeweisen und Modellgenerierung) im Normativen Schließen; • Agentenbasierte Simulationsplattform zum Experimentieren mit ethischen und rechtlichen Theorien; • Fallstudien mit dieser Simulationsumgebung zur Eignungsbeurteilung ethischer und rechtlicher Theorien; • Entwurf und Modellierung von Architekturen für "Ethical Governors"; • Exemplarische Implementierungen, Erprobung und Bewertung solcher Architekturen in ausgewählten Bereichen: (i) Autonome Fahrzeuge (mit Rojas und Göhring an der FU Berlin), (ii) Finanz- und Bankwesen (mit KPMG Lighthouse Luxembourg), (iii) Soziale Robotik (mit LuxAI S.A.), (vi) Pharmazie and Gesundheitswesen (mit Latentine GmbH), (v) Kognitive Robotik (mit DLR), and (vi) KI-basiertes Marketing (mit wizAI solutions GmbH); • Konferenzen, Workshops, und Tutorien zu den Themen von ENoRME • Vorlesungen, Seminare, e-Learning-Ressourcen, Promotions- and Studentenprojekte zu den Themen von ENoRME; • Öffentlichkeitsarbeit.

Perspektiven einer Verstetigung der Kooperation: ENoRME wird bereits bestehende Kooperation mit der U Luxembourg, KPMG Luxembourg und der U Miami ausbauen zu einem internationalen Netzwerk mit gemeinsamer, nachhaltiger Infrastruktur: Kernkomponente wird die im Projekt entstehende, leicht verfügbare, cloud-basierte Werkbank für Entwicklungs- und Anwendungsstudien in der Maschinen-Ethik sein. Die Nachhaltigkeit der Kooperation in diesem Netzwerk soll zudem gefördert werden durch die Akquise weiterer Projekte über den beantragten Projektzeitraum hinaus. Weitere Synergieeffekte sollen geschaffen werden, z.B. durch Kooperationen mit: • Berlin Mathematics Research Center MATH+• der Doktorandenschule *"Power, Humanity, and Technology: A History of Desire and Control to Inform the Future of AI"* (Antragstellung erfolgt im Herbst), • EU MIREL Network • anderen Arbeitsgruppen, Institutionen und verwandten Initiativen im Bereich Verantwortungsvolle KI im Berliner und Luxemburger Umfeld. ENoRME wird auch den Wissenstranfer mit dem Industriesektor in diesem wichtigen Gebiet nachhaltig fördern, durch Kooperation mit Mittelstand 4.0-Kompetenzzentren und mit thematisch verwandten Projekten im Rahmen weiterer Bundes- und Europa-Initiativen (z.B. CLAIRE [32, 87] and Digital SME Alliance).

Öffentlichkeitswirksame Kommunikationsmaßnahmen: Zielgruppenspezifische Kanäle für: allgemeines Publikum (Facebook, Twitter, Science Slams), Fachleute (LinkedIn, Xing), Forscher/innen (ResearchGate, Academia.edu); Content-Marketing (Fachzeitschriften, Fachblogs, IT-Webseiten, etc.); Public AI Lab; Moderne Webpräsenz.

Geschätzte Ausgaben/Kosten

Gesamt	ca. €4.990.000

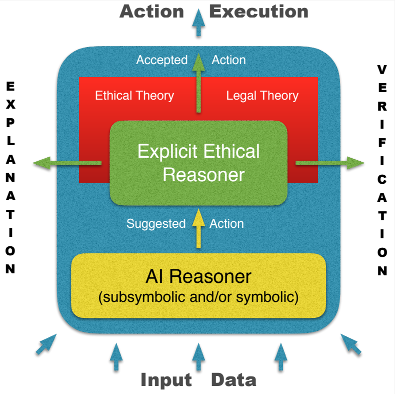
[a] In Klammern werden führende Wissenschaftler/innen der partizipierenden Institutionen genannt; mit diesen wird ein nachhaltiges Netzwerk aufgebaut—zusätzlich zu den entsendeten direkt partizipierenden Wissenschaftler/innen.

1 Motivation and Core Objectives for ENoRME

Intelligent autonomous systems (IASs) are rapidly entering applications in industry, military, finance, governance, administration, healthcare, etc., leading to an historical transition period with unprecedented dynamics of innovation and change, and with unpredictable outcomes; see, e.g., Floridi [43], Tegmark [95], Reese [80] and the references therein. Politics, regulatory bodies, indeed society as a whole, are challenged not only with keeping pace with these potentially disruptive developments, but also with staying ahead and wisely guiding the transition. Some scientists are even concerned about an emerging superintelligence, a vision that we don't share, but that we cannot rule out. Since there is much at stake, preventive investments in the exploration, development and implementation of appropriate means of controlling IASs are absolutely justified. **A balanced approach to AI is needed**, fostering positive impacts, while preventing negative side effects. This vision is shared, e.g., in the "Ethics Guidelines for Trustworthy AI" [41] of the European Commission's High-Level Expert Group on AI, the German AI strategy [96] and the OECD recommendations [74].

Many AI researchers, developers, and users are concerned with challenges such as "How to stay ahead . . . ", "How to keep up with . . . ", and "How to close the gap to . . . " the rapid developments in AI. A global technology race is the result. A much smaller number of AI researchers, developers, and users are concerned with researching and developing means of wisely regulating and controlling future IASs. **ENoRME addresses these issues.** The primary focus, however, is on risk prevention, more precisely, on explicit ethical governance of IASs [2, 3, 83].[1] We see it as a societal mandate to invest in risk preventing technology; such a sensitive issue should not be driven by commercial interests alone. Despite of the **focus on "Trustworthy AI"**, the range of applications that will be enabled by ENoRME is wide and colourful.

ENoRME contributes foundational technology to foster, aside from many other topical applications, the **development of legal and ethical governors** for IASs. These governors will provide much needed symbolic, deductive means of control, which operate orthogonally to, and in addition to, corresponding mechanisms at the less transparent, less accountable, and less trustworthy level of non-symbolic reasoning. ENoRME will contribute a much needed, powerful and flexible, on-demand **Universal Reasoning Workbench (URW), with a particular emphasis on Normative Reasoning**. It will explore and demonstrate the utilisation of this URW in selected **case studies and experiments**.



European Commission – Press Release (Brussels, April 8, 2019):

Artificial Intelligence: Commission takes forward its work on ethics guidelines

Andrus Ansip (VP for the Digital Single Market): *"The ethical dimension of AI is not a luxury feature or an add-on. It is only with trust that our society can fully benefit from technologies. Ethical AI is a win-win proposition that can become a competitive advantage for Europe: being a leader of human-centric AI that people can trust."*

Seven essentials are mentioned in this press release (see also [41]) for achieving trustworthy AI:
A Human agency and oversight, **B** Robustness and safety, **C** Privacy and data governance, **D** Transparency, **E** Diversity, non-discrimination and fairness, **F** Societal and environmental well-being, **G** Accountability

The core objectives of ENoRME address these essentials (in particular, **A**, **B**, **D** and **G**). The overall challenge is of interdisciplinary nature; **ENoRME therefore bundles and integrates interdisciplinary expertise**: Computer Science (CS) and Engineering, Robotics and Machine Learning (ML), Logic and Deduction Systems, Philosophy, Formal Ethics. The research methods in ENoRME range over a wide spectrum: foundations and theory, design, modelling, implementation, formal verification, and empirical assessment. ENoRME will demonstrate how the novel reasoning tools and developed infrastructure can be utilised in practice for controlling the behaviour of IASs, with explicit legal and ethical theories. Moreover, all of the developed technology will be integrated into an on-demand, cloud-based Universal Reasoning Workbench (URW) in order to foster knowledge transfer with industry, research, and educational institutions, and particularly to enable their **access to critical AI infrastructure at scale with little risk and minimal costs**.

[1] Floridi [43, p.13]: *". . . the challenge ahead will not be so much digital innovation per se, but the governance of the digital, AI included."*

2 Location and Coordinating Institution of ENoRME

Location: Berlin-Dahlem. ENoRME will benefit from proximity to politics, government, media, a vibrant start-up scene, and an excellent academic environment. Berlin is particularly attractive for guest researchers, and for the organisation of international workshops and conferences on the topic areas of ENoRME.

Coordinating Institution: Freie Universität Berlin, Department of Mathematics and Computer Science (CS), Dahlem Center for Machine Learning and Robotics (DCMLR). FU Berlin is one of Germany's Excellence-Universities; ENoRME will foster and benefit from local collaborations and infrastructure (e.g. FU Berlin has a very active press office).

Strategic Integration: FU Berlin will benefit from ENoRME in multiple ways: (i) it will help preventing an eventual brain drain in a strategically important area for the university and for Berlin (Benzmüller and Göhring are not holding tenured professorship positions); (ii) it will recruit further expertise in areas not yet represented at FU Berlin, and this will be beneficial for education, training and project acquisition; (iii) mutual fertilisation will be fostered with other ongoing projects, such as the Berlin Mathematics Research Center MATH+, research in the secure and safe deployment of computer systems, and with planned projects (e.g. project/doctoral school on History & AI with JFK-Institute).

Room Allocation: In the initial few months it will be possible to host about 6-8 ENoRME-researchers at the DCMLR. In the second and third year, it will eventually be required to rent office space nearby. This has already been discussed with the Dep. of Mathematics and CS of FU Berlin. We have allocated a room rent (€350,000) in the budget.

3 Composition of the ENoRME International Consortium[2]

Participating Partner Nodes. Germany: Freie Universität Berlin, University of Koblenz-Landau; **EU**: University of Luxembourg (LU), University of Bergen (NO); **Other**: University of Miami (US), University of Liverpool (UK[3]), University of Campinas (BR), University of Buenos Aires (AR)

Further Collaborators. Industry/Institutes: Latentine GmbH (SME[4], DE), wizAI solutions GmbH (SME, DE), LuxAI S.A. (SME, LU), KPMG Lighthouse Luxembourg (LU), DLR (DE); **Academia:** DHBW Stuttgart (DE), U Utrecht (NL), UNSW/Data61 (AU), Zhejiang University (CN), FU Berlin (DE), Gesellschaft für Informatik e.V. (DE), The American University of Paris (FR), Swansea University (UK), Infosys (US).

In addition to the research collaborations, a teaching collaboration with an interdisciplinary project/doctoral school on *"Power, Humanity, and Technology: A History of Desire and Control to Inform the Future of AI"* (History & AI; proposal to be submitted shortly); PI's besides Benzmüller and Göhring are Prof. Gienow-Hecht (JFK-Institute), and Prof. Johnston (U of Ottawa). If accepted, we expect a significant knowledge transfer from ENoRME to the JFK-Institute-project. ENoRME may also benefit from this interaction but is not depending on it. ENoRME will also interact with MATH+, which, among others, studies the *mathematics of machine learning*. Potential further collaborator: Prof. Rahwan, Center for Humans and Machines, Max-Planck-Institute for Human Development in Berlin.

3.1 Participating Researchers and their Specific Expertise for ENoRME

An overview on the participating researchers is in provided in Figure 1. We here provide further details, including selected relevant publications:

Freie Universität Berlin, Dahlem Center for Machine Learning and Robotics

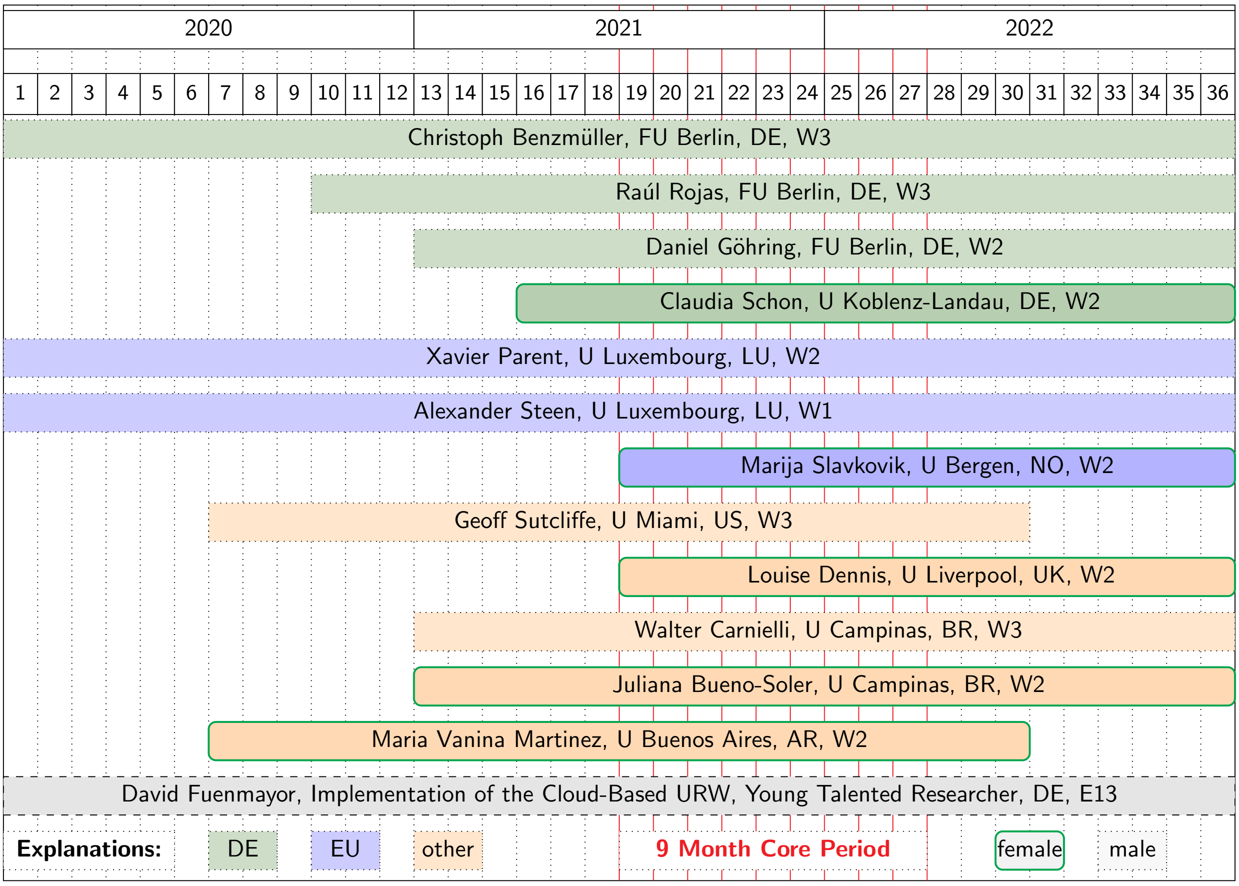
1. **Christoph Benzmüller**, Prof. Dr. habil. [Url] [E-Mail]; **Participation:** 01/20–12/22; 36 months (W3); **Expertise:** Universal Logic Reasoning, Automated Theorem Proving (ATP), Knowledge Representation & Reasoning (KR), Formal Methods; **Education:** Dipl. in CS (1995), PhD in AI (1999), Habil. in CS (2007), Full Professorship (2008), Venia Legendi in CS & Maths (2012); **Highlights:** DFG Heisenberg fellowship (2012-2017), VP of CADE, steering committee of DEON, many related keynotes; **Publications:**[5] [5], [8]*, [9]*; **Notes:** member of CLAIRE, experienced organiser/chair of AI conferences, co-PI of related project/doctoral school proposal with JFK-Institute.

[2] The CVs and letters of intent of all partners of ENoRME are attached. Important remark: Title and acronym of the project have changed last minute; some attached letters of intent or CVs may thus still use an earlier title and/or acronym; we apologise.

[3] Brexit is scheduled for 2019, we thus count UK as a non-EU partner site here.

[4] Small or Medium size Enterprise; same as the German expression "KMU".

[5] Joint publications with other ENoRME-researchers are marked with *; they result from existing bilateral collaborations.



Collaboration with Industry: Freek Stulp (DLR, Oberpfaffenhofen, DE), Sven Mühlenbrock (KPMG Lighthouse, LU), Ulrich Furbach (wizAI solutions GmbH, SME, DE), Pouyan Ziafati & Aida Nazarikhorram (LuxAI, SME, Luxembourg, LU), Max Wisniewski (Latentine GmbH, SME, Berlin, DE)

Further Academic Collaborators: Jan Broersen (U Utrecht, NL), Michael Fisher (U Liverpool, UK), Beishui Liao (Zhejiang U, CN), Daniel Krupka (Gesellschaft für Informatik e.V., DE), Tomer Libal (American U of Paris, FR), Adam Pease (Infosys, US), Ralf Romeike (FU Berlin, DE), Stephan Schulz (DHBW Stuttgart, DE), Leon van der Torre (U Luxembourg, LU), Toby Walsh (UNSW Sydney/Data61, AU), Adam Wyner (Swansea U, UK),

Figure 1: ENoRME person months distribution of participating researchers, and list of further collaborators

2. **Raúl Rojas**, Prof. Dr. Dr. (h.c.) habil. [Url] [E-Mail]; **Participation:** 10/20–12/22; 27 months (W3); **Expertise:** KI, Robotics, Autonomous driving, ML; **Education:** MSc in Maths (1980), Master of Economics (1981), PhD in Economics and Social Sciences (1988), Habil. in CS (1994), Doctorate **honoris causa** in CS (2016); **Highlights:** Professor of the Year 2014 in Germany, National Science Prize 2015 of Mexico, Member of the Mexican Academy of Sciences; **Publications:** [81] [28] [33].
3. **Daniel Göhring**, Prof. Dr. [Url] [E-Mail]; **Participation:** 01/21–12/22; 24 months (W2); **Expertise:** Robotics, Autonomous driving, Cooperative behavioral planning, ML; **Education:** Diploma in CS (2004), PhD in Robotics/AI (2009), W1-Prof. in Robotics (2015); **Highlights:** Founder and Head of Autonomous Car Labs in Berlin; **Publications** [61] [59]* [60]; **Notes:** co-PI of related project/doctoral school proposal with JFK-Institute.

Universität Koblenz-Landau, Institute for Web Science and Technologies

4. **Claudia Schon**, Dr. [Url] [E-Mail]; **Participation:** 04/21–12/22; 21 months (W2); **Expertise:** Description logics, KR, Commonsense reasoning, ML, Ontologies.; **Education:** Trained inform. techn. officer (2000), Dipl. in CS (2006), PhD in KR (2016); **Highlights:** WS-Organizer "Bridging the Gap between Human and Automated Reasoning", PI of DFG project on "Cognitive Reasoning", Co-speaker of the Special Interest Group for Deductive Systems of the Gesellschaft für Informatik e.V. (GI), Klara Marie Faßbinder Guest Professorship (2019); **Publications** [85] [52, 53]*.

University of Luxembourg, Interdisciplinary Lab for Intelligent and Adaptive Systems

5. **Xavier Parent**, Dr. [Url] [E-Mail]; **Participation:** 1/20–12/22; 36 months (W2); **Expertise:** KR and AI, Normative Multi-Agent Systems (MAS), Deontic logic; **Education:** PhD in Philosophy (2002); **Highlights:** Co-editor of

Handbook of Deontic Logic and Normative Systems and of textbook series *Texts in Logic and Reasoning*; invited speaker/tutorial at DEON'16, SSLPS'19, KI 2019; **Publications** [56] [76] [77] **Notes:** Member of CLAIRE.

6. **Alexander Steen**, Dr. [Url] [E-Mail]; **Participation:** 1/20–12/22; 36 months (W1); **Expertise:** ATP, Deduction Systems, System Development (Leo-III); **Education:** BSc in Maths (2014), BSc (2012) and MSc (2014) in CS, PhD in AI (2018); **Highlights:** Junior-Fellow of the German Informatics Society (GI), speaker of the Special Interest Group for Deductive Systems of GI, board member of AI chapter of GI; **Publications** [58]* [89]* [88].

University of Bergen, Logic, Information and Interaction Group

7. **Marija Slavkovik**, Prof. Dr. [Url] [E-Mail]; **Participation:** 7/21–12/22; 18 months (W2); **Expertise:** Collective Reasoning, Machine Ethics, MAS; **Education:** Dipl. (2005), MSc in CS (2007), PhD in Logic (2012); **Highlights:** VP of the Norwegian AI Association, assoc. editor of AI Review, member of advisory group on Ethics, Legal, Social Issues (ELS) of CLAIRE; co-editor special issue on Machine Ethics of the AI Journal; **Publications** [69]* [35]* [66].

University of Miami, Department of Computer Science

8. **Geoff Sutcliffe**, Prof. Dr. [Url] [E-Mail]; **Participation:** 7/20–06/22; 24 months (W3); **Expertise:** Automated reasoning, Practical applications & evaluation of reasoning systems, System development, Standardisation; **Education:** BSc in CS & Maths (1983), MSc in CS (1986), PhD in CS (1992), Full Professor of CS (since 2014); **Highlights:** Creator of the TPTP problem library, Founder and host of the annual CASC championships, TPTP language standardization for ATP systems; **Publications:** [90] [94] [92]*.

University of Liverpool, Autonomy and Verification Laboratory

9. **Louise Dennis**, Dr., Lecturer [Url] [E-Mail]; **Participation:** 07/21–12/22; 18 months (W2); **Expertise:** Autonomous Systems, MAS, ATP, System Development, Machine Ethics; **Education:** BSc in Maths & Phil. (1992), MSc in CS (1994) PhD in AI (1998), Postgraduate Certificate in Higher Education (2005); **Highlights:** Member of IEEE Global Initiative for Ethical Considerations in the Design of Autonomous Systems (sub-committee Embedding Values into Autonomous Intelligent Systems); **Publications** [35]* [36] [42].

State University of Campinas, Centre for Logic, Epistemology and the History of Science

10. **Walter Carnielli**, Prof. Dr. [Url] [E-Mail]; **Participation:** 1/21–12/22; 24 months (W3); **Expertise:** Non-Classical Logics, Logic Combinations, Semantics, Proof-Theory, Computability Theory incl. Quantum Computing, Reasoning with Inconsistent & Incomplete Inform., Argumentation, Phil. of Logic; **Education:** PhD in Maths (1982); Habil. in Logic & Foundations of Maths (1990); **Highlights:** Research Fellow with L. Henkin at UC Berkeley, A.v.Humboldt Stip. U Münster; Member of Brazilian Advanced Institute for AI (AI2); **Publications:** [30] [29] [31].
11. **Juliana Bueno-Soler**, Prof. Dr. [Url] [E-Mail]; **Participation:** 1/21–12/22; 24 months (W2); **Expertise:** Non-Classical & Modal Logics, Algebraizability of Logics, Found. of Probability; **Education:** BSc in Maths (2000), MSc (2004) and PhD in Phil. (2009); **Highlights:** VP of Brazilian Society of Logic (SBL); **Publications:** [25] [24] [27].

University of Buenos Aires, Department of Computer Science

12. **Maria Vanina Martinez**, Prof. Dr. [Url] [E-Mail]; **Participation:** 07/20–06/22; 24 months (W2); **Expertise:** Reasoning under Uncertainty, Semantic Web, Question Answering & Argumentation, Inconsistency Management, Defeasible Reasoning; **Education:** Dipl. in CS (2005), PhD in CS (2011); **Highlights:** IEEE Intelligent Systems AI's Ten to Watch (2018), IJCAI Early career Spotlight (2017); **Publications** [57, 72, 70].

Technical Design and Implementation of the Cloud-Based URW, Young Talented Researcher

- **David Fuenmayor** [Url] [E-Mail]; **Participation:** 1/20–12/22; 36 months (E13); **Expertise:** Software Engineering, Cloud-based systems, Logic in AI, Argumentation, NL semantics; **Education:** Engineer Degree (2009) and MSc in Mechatronics (2012), BSc in Philosophy (2018), current PhD candidate in AI/Logic **Highlights:** Professional software developer and architect (cloud technologies), ≥ 5 years of work experience in enterprise software projects (IoT, automotive and energy sector); **Publications** [50]*, [49]*, [47]*; **Responsibilities:** (primary) technical design & implementation of URW, (secondary) research and assisting the coordination of ENoRME.

3.2 Further Collaborators of ENoRME

Industry and National Research Institutes. ENoRME will collaborate with industry and other research institutes in selected case studies. Each of these case studies will be conducted with the help of student researchers and supervised jointly by two of the participating researchers and the external partners. The external partners will provide expertise, access to knowledge, and infrastructure as required.

1. **DLR**, Department of Cognitive Robotics – Institute of Robotics and Mechatronics, Oberpfaffenhofen **Contact:** Freek Stulp, Prof. Dr., Head of department [Url] [E-Mail]; **Description:** Applications of AI and ML to the control and planning of robots; **ENoRME-Partner:** Benzmüller, Göhring, Rojas.
2. **KPMG Lighthouse Luxembourg**; **Contact:** Sven Mühlenbrock, Dr., lead, [Url] [E-Mail]; **Description:** Applications of AI and data analytics in banking and finance; **ENoRME-Partner:** Benzmüller, Parent, Steen.
3. **Latentine GmbH**, SME in Berlin; **Contact:** Max Wisniewski, MSc, CTO, [Url] [E-Mail]; **Description:** Industrial applications of ML and symbolic AI e.g. in pharmacy and healthcare; **ENoRME-Partner:** Steen, Sutcliffe.
4. **LuxAI S.A.**, SME/spin-off of U Luxembourg, **Contact:** Pouyan Ziafati, Dr. & Aida Nazarikhorram, Dr., [Url] [E-Mail]; **Description:** Applications of AI in Social Robotics; **ENoRME-Partner:** Benzmüller, Parent.
5. **wizAI solutions GmbH**, SME in Koblenz; **Contact:** Ulrich Furbach, Prof. Dr. [Url] [E-Mail]; **Description:** Digitalisation and AI; **ENoRME-Partner:** Schon, Steen.

International Network of Experienced Researchers. The ENoRME project will create an international network beyond the above mentioned participating researchers. In particular, the following international leading scientists will contribute their expertise to ENoRME on the basis of shorter research visits, participation in workshops, tutorials and conferences organised by ENoRME and by co-supervising and critically reviewing our research efforts, etc. Some of these additional researchers are affiliated with the network nodes mentioned above. A particular emphasis is on laying the foundations for a strong and sustainable international research network on the topics of ENoRME in combination with early dissemination and outreach activities. Joint supervision of student projects is a also a major objective. A particular focus is on establishing a sustainable collaboration between FU Berlin and U Luxembourg.

6. **Jan Broersen**, Prof. Dr. (U Utrecht, NL) [Url] [E-Mail]; **Expertise:** Architectures for ethical governors, Formal ethics, Normative reasoning, Logics of agency; **Highlights:** ERC consolidator grant *"Responsible Intelligent Systems (REINS)"*, President of DEON.
7. **Michael Fisher**, Prof. Dr. (U Liverpool, UK) [Url] [E-Mail]; **Expertise:** Autonomous systems technology, Verification & Validation; **Highlights:** Head of UK Network on the Verification & Validation of Autonomous Systems, Fellow of BCS and IET, Member of the BSI AMT/10 committee on Robotics and the IEEE P7009 Standard for Fail-Safe Design of Autonomous Systems.
8. **Daniel Krupka**, Gesellschaft für Informatik e.V., Berlin [Url] [E-Mail]; **Expertise:** Science Communication.
9. **Beishui Liao**, Prof. Dr. (Zhejiang U, CN) [Url] [E-Mail]; **Expertise:** Reasoning with uncertain and incomplete information, Argumentation, MAS, Legal and ethical AI; **Highlights:** Steering committee of DEON, organiser of *"Chinese Conference on Logic and Argumentation (CLAR)"*.
10. **Tomer Libal**, Prof. Dr. (American U in Paris, FR) [Url] [E-Mail]; **Expertise:** Legal informatics, Proof transformation.
11. **Adam Pease**, Dr. (Infosys, Palo Alto, US) [Url] [E-Mail]; **Expertise:** Deep semantic NLP, Upper ontologies.
12. **Ralf Romeike**, Prof. Dr. (FU Berlin) [Url] [E-Mail]; **Expertise:** Didactics of CS, Education tools.
13. **Stephan Schulz**, Prof. Dr. (DHBW Stuttgart, DE) [Url] [E-Mail]; **Expertise:** ATP, ML, System development; **Highlights:** Developer of E (leading first-order ATP), Multiple co-chair of *"Conf. on AI and Theorem Proving"*.
14. **Leon van der Torre**, Prof. Dr. (U Luxembourg, LU) [Url] [E-Mail]; **Expertise:** Normative reasoning, Deontic logic, MAS, Cognitive robotics, Agreement technologies; **Highlights:** PI of EU funded research network MIREL: MIning and REasoning with Legal texts, Editor of the Deontic Logic Corner of the Journal of Logic and Computation, Co-editor of Handbook of Deontic Logics and Normative Reasoning.
15. **Toby Walsh**, Prof. Dr. (UNSW Sydney/Data61 & TU Berlin, AU) [Url] [E-Mail]; **Expertise:** Trustworthy AI, Algorithmic decision theory; **Highlights:** Fellow of Australian Academy of Science, Fellow of the Association for the Advancement of AI, Fellow of the European Association for AI, Enormous media presence on Trustworthy AI.
16. **Adam Wyner**, Prof. Dr. (Swansea U, UK) [Url] [E-Mail]; **Expertise:** AI & Law, Legal Informatics. **Highlights:** PC chair JURIX 2017 & ICAIL 2021, Director of Centre on Innovation and Entrepreneurship in Law

4 Research and Development Plan, with Justification of Consortium

Progress in ML, in particular deep learning, has enabled impressive recent success stories in the development and deployment of IASs. In ENoRME we will target pressing challenges that these successes have generated. A major concern is that IASs that rely exclusively on non-symbolic technologies will increasingly lack transparency, explainability, verifiability, and ethical behaviour – the essential requirements for "Trustworthy AI" [74, 41, 96].[6] **A particular challenge concerns the development of mechanisms of ethical and legal control for future IASs**, such as the ability to construct and reason with high-level conceptual representations of legal and ethical concepts. **A convincing solution must fruitfully integrate non-symbolic with symbolic techniques, based on explicit knowledge representation and reasoning**. It is particularly the latter where Europe and Germany can capitalise and further expand its existing leading expertise.[7]

The success of deep learning methods is based on the availability of huge amounts of data, and relies on the impressive growth in hardware capabilities, allowing for massively parallel data processing. In particular, the adoption of general-purpose graphical processors (GPUs) has been a key technological enabler. More recently, other kinds of hardware architectures have increasingly been utilized to further accelerate ML training and inference tasks. In particular, the utilisation of reconfigurable hardware approaches (particularly Field Programmable Gate Arrays) has experienced a renaissance. Other big-data technologies, e.g., graph databases and the semantic web, have also taken advantage of massively parallel processing on server farms combining multi-core CPUs and GPUs. **The area of automated reasoning in expressive logics is, however, lagging behind in this regard. A building block that is missing on the symbolic side is powerful automated reasoning technology for expressive non-classical logics and their combinations, as required for realising adequate forms of automated and semi-automated normative reasoning.** The development of competitive automated reasoning technology for such ambitious logics typically requires substantial resource investment and expertise, which are typically not available to small research teams or SMEs. It is clear that any strategy aimed at **democratising access to AI**, in particular for SMEs and educational institutions, must facilitate their access to such critical hardware resources. **Developing, sharing, and standardising such technology to stimulate research and deployment is one important mission of ENoRME.**

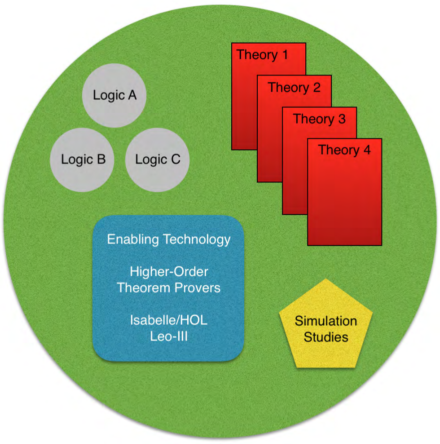
The ENoRME R&D vision has been specifically geared towards simultaneously addressing these two concerns: **developing mechanisms of legal and ethical control, and providing the necessary reasoning infrastructure for that reasoning.** In addition to boosting knowledge transfer of our research outcomes with the scientific community, industry, and the public sector, ENoRME will address the real need of SMEs and educational institutions for **experimenting with AI at scale, with little risk and low cost.** A primary focus thereby is on the provision of technology for automating expressive normative reasoning. This is beneficial for AI research in a wide sense, not only for the legal and ethical governing of IASs. **ENoRME will thus develop, as its core deliverable, an integrated cloud-based, on demand, Universal Reasoning Workbench (URW) with a specific focus on normative reasoning**. This URW will support modelling and automated experimentation with explicit and legal theories for use in ethical governors within IASs. The URW will also find application in other areas, e.g., rational argumentation, natural language processing, computational metaphysics, etc. Moreover, all of ENoRME's solutions will be applicable both independently and in combination with alternative means based on ML and other non-symbolic approaches. **The overall objective is to contribute a missing explicit reasoning component in the development of secure, explainable, and trustworthy AI systems.**

At the same time as providing pragmatic solutions, ENoRME will engage in further theoretical and experimental research in explicit ethical governing mechanisms for IASs. **In this sense ENoRME will become its own first customer.** The idea is to create an immediate **feedback loop** between (i) the URW, (ii) the ethical governing architectures that we will devise and adopt, and (iii) their experimental application in selected case studies with our industrial partners. This feedback loop will then particularly inform our work on the URW.

[6] We acknowledge the valuable efforts being made inside the deep learning research community towards increasingly enabling the extraction of higher-level representations from modularised neural networks. However, there are still many open research questions. Regarding interpretability of outcomes (cf. "explainable AI"), such mechanisms are rather laborious and only useful in a kind of 'forensic' way. AI systems need explicit, 'on-line' interpretability.

[7] Moreover, research on the integration of non-symbolic and symbolic AI is relevant much beyond the specific application focus of ENoRME: it is at their melting point where the "next big thing" is to be expected and where challenging questions, also regarding the emergence of a superintelligence, can be tackled and eventually answered.

The **Starting Point** for ENoRME and the URW will be recent results in **universal logical reasoning** (cf. [5, 6] and the references therein) that are rooted in Benzmüller's research group. These results have been picked-up in various bilateral collaborations [13, 62, 15]. ENoRME will create a momentum in which such collaboration activities, in particular in the area of normative reasoning, will be bundled in a most resource-effective and impactful manner. Since this challenge is of an interdisciplinary nature, **the ENoRME network will consolidate interdisciplinary expertise**. To initiate practical deployment and assessment of research outcomes, ENoRME will integrate **high-impact interaction with industry** (Latentine GmbH, LuxAI S.A., KPMG Lighthouse Luxembourg, wizAI solutions GmbH) and research institutes (DLR); nearly all of these collaborations are founded on prior bilateral interactions. By leveraging our cloud-based URW, ENoRME will develop and offer **customized training and qualification programs for students, researchers, and engineers**. Among these efforts ENoRME will offer courses, seminars, conferences, workshops, and e-Learning resources (including a MOOC).



Core Components of our URW are depicted graphically in the picture on the right. The framework supports experimentation with different normative theories, in different application scenarios, and is not tied to a specific ethical theory. One enabling technology will be higher-order reasoning systems such as Leo-III [89] and Isabelle/HOL [73], via the universal logical reasoning approach [5, 6]. The figure displays different components and aspects that are relevant. Different target logics (grey circles) and their combinations are provided in the meta-logic of the host system. Different ethical and legal theories can then be encoded in the target logics. This enables two-way fertilisation: In one direction the theories can be assessed for, e.g., consistency, entailed knowledge, etc. In the other direction properties of the different target logics can be investigated. For example, while one target logic might suffer from paradoxes in a concrete application context, another target logic might well be sufficiently stable against paradoxes, and also practically responsive in the same application context. Suitable logic-theory pairings will be further assessed in the project: (i) in form of simulation studies, e.g., within simulated agent societies, where the single agents are constrained by the theory, and (ii) in concrete, real world experiments, e.g., within autonomous cars or social robots.

The **Objectives and Deliverables** of ENoRME comprise: • A cloud-based, on demand, URW, including an associated experimentation platform; • Standardisation and benchmarking in normative reasoning; • Theory and implementation of novel, relevant logic combinations; • Examples of mechanised and automated ethical and legal theories; • Additional specialist ATP and model finding technology for expressive normative reasoning; • An agent-based simulation environment enabling experiments with ethical and legal theories; • Case studies conducted within this agent-based simulation environment to assess different legal and ethical theories; • Design and modelling of ethical governor architectures; • Exemplary implementation, experimentation, and assessment of ethical governor architectures equipped with legal and ethical theories, including (i) autonomous cars (with Rojas and Göhring at FU Berlin), (ii) finance and banking (with KPMG Lighthouse Luxembourg), (iii) social robotics (with LuxAI S.A.), (vi) pharmacy and healthcare (with Latentine GmbH), (v) cognitive robotics (with DLR), and (vi) AI-driven marketing (with wizAI solutions GmbH). • Conferences, workshops, and tutorials on the topics of ENoRME • Lecture courses, seminars, e-Learning resources, PhD and student projects on the topic of ENoRME • Outreach: public debates, media appearances, science slams, website, Facebook, Twitter, etc.

Relationship with European Initiatives. Widespread social acceptance of AI is essential for leveraging both its benefits in society and for the competitive advantages its wise application can produce in the global economic and technological landscape. ENoRME will contribute technology that will help fulfil three main requirements (as put forward by HLEG [41]) that any **Trustworthy AI** should meet throughout the entire system's life cycle: (i) it should be lawful, complying with all applicable laws and regulations; (ii) it should be ethical, ensuring adherence to ethical principles and values; and (iii) it should be robust, both from a technical and social perspective, since, even with good intentions, AI systems can cause unintentional harm. The implementation of our ethical approach towards AI is not at odds with the growing need for powerful data processing infrastructure. For instance, a common criticism of data protection initiatives concerns the (alleged) hampered access to the massive amounts of training data needed for current deep learning techniques. In contrast, he ENoRME approach to trustworthy AI, having a particular emphasis on explicit logical representations, features the characteristic of enabling reasoning with high-level conceptual abstractions, thus trading quantity for quality of data. The technological challenges set by European data protection initiative indeed provide us with an opportunity to take the lead in designing better AI systems, able to provide explanations for decisions that significantly affect individual rights and liberties, and enabling effective auditing mechanisms to ensure unbiased and discrimination-free decision making.

State of the Art, Prior Research and Added Value of ENoRME. The question of how transparency, explainability, and verifiability can best be achieved in future IASs, and whether bottom-up or top-down architectures should be preferred, are discussed in various articles, e.g., [40, 100, 38, 83, 71, 1, 99, 35]. For example, Dennis *et al.* [35] make a compelling case for the use of so-called formal verification – a well-established technique for proving correctness of computer systems – in the area of machine ethics. ENoRME develops a reasoning infrastructure for such normative reasoning. Such an infrastructure is much needed and related work is thin.

Members of the ENoRME consortium have made various contributions of research and development in topics related to relevant areas of ENoRME. These are surveyed next.

Research and concrete results for the automation of reasoning for quantified deontic logics, and their combinations with other non-classical logics (epistemic, temporal, defeasible, probabilistic, etc.) as required in realistic applications of normative reasoning, is sparse. A connection-based ATP covering among others, first-order standard deontic logic, has been developed by Otten [75]. Schon and Furbach [54] utilise the tableaux-based Hyper prover [4] for first-order logic to automate deontic reasoning. First-order resolution methods for propositional modal logics have been contributed by Schmidt and Hustadt [84]. Further related work includes a reasoner for propositional defeasible modal logic by Governatori and his team [63], a reasoner for expressive contextual deontic reasoning by Bringsjord *et al.* [20], and Prolog based implementations of deontic and counterfactual reasoning by Pereira and Saptawijaya [78, 82, 79].

Benzmüller, Parent, and van der Torre have collaborated intensively since 2016 on the automation of propositional and quantified deontic logics, on the development of an overall methodology for the automation of normative reasoning, and on the use of their approach in some small, selected case studies [13, 8, 10, 9, 12]. A larger case study in formal ethics, which required the integration of a higher-order dyadic deontic logic with an appropriate notion of context, was subsequently contributed by Fuenmayor and Benzmüller [50, 48]. The approach integrates the results of different research communities: existing state-of-the-art reasoning tools for classical higher-order logic, such as Leo-III and Isabelle/HOL, which internally cooperate again with state-of-the-art first-order ATPs and SMT solvers, are adapted and reused to support normative, non-classical reasoning. The enabling methodology is universal (meta-)logical reasoning [5, 6] in classical higher-order logic [7].

U Luxembourg and FU Berlin have collaborated on topics related to responsible AI since 2016. They have organised joint lecture courses, joint student supervision, joint conferences, etc. An example is the LuxLogAI 2018 summit, with more than 200 participants in 11 subevents related to AI. It included conferences such as RuleML+RR [11] and GCAI [67], and held a very successful public round table discussion focused on Responsible AI and the European future, in collaboration with KPMG Luxembourg. This existing collaboration structure will be further expanded in ENoRME to form a strategic alliance on Trustworthy AI.

Steen and Benzmüller developed Leo-III [89, 88], a leading[8] ATP system for classical higher-order logic [58]. Leo-III has been adapted to pioneer the automation of more than 120 different quantified (multi-)modal logics [58], and also for an initial range of quantified deontic logics.

Benzmüller and Sutcliffe developed the TPTP THF standards for classical higher-order logic [92, 93, 14]. Funded by a grant from the European Commission,[9] this work has initiated expansion of Sutcliffe's international theorem proving competition CASC [91] to higher-order logic, which in turn fostered the current shift of interest in the community from first-order ATP to higher-order ATP – leading ATP systems such as Vampire [64], E [98, 86], and CVC4 [37] are currently being extended in to higher-order logic.

Rojas and Göhring are experts in robotics, autonomous driving, and ML. They head the AUTONOMOS Labs of FU Berlin, and have collaborated with Benzmüller at the DCLMR since 2011. ENoRME will provide them with the

[8] An independent recent study has confirmed the excellent performance of this system [23].

[9] See project report at https://cordis.europa.eu/project/rcn/88314/reporting/en

resources and expertise necessary for experimentally linking their research in robotics and autonomous driving to explicit ethical governing architectures.

Slavkovik, Liao, and van der Torre have collaborated on a methodological level to develop an artificial moral agent architecture that uses techniques from normative systems and formal argumentation to reach moral agreements among stakeholders [39, 68]. Slavkovik is co-editing a special issue on "Ethics for Autonomous Systems" to appear in the flagship journal "Artificial Intelligence".

Dennis and Fisher have studied ethically critical machine reasoning and have proposed practical architectures and reasoning tools [19, 34] that we will integrate with the ENoRME-infrastructure in order to stimulate mutual fertilisation. Dennis has also worked with Slavkovik on formalisation and verification of ethical governors [35], construction of benchmarks for ethical reasoning [18], formal reasoning about privacy in social networks and digital crowds.

Schon and Furbach are particularly interested in modelling human reasoning within automated reasoning systems, and in reducing the gap between these antipodes [51, 55, 97, 65]. A particular emphasis has been on modelling affective aspects of human reasoning. Schon is also an expert in question answering systems and in commonsense reasoning.

Martinez adds relevant further expertise in knowledge engineering, ontologies, semantic web, reasoning under uncertainty and question answering. She participates as a researcher in the MIREL project (coordinated by U Luxembourg).

Carnielli and Bueno-Soler are leading experts in paraconsistent reasoning and combinations of logics [26, 30, 29, 17]. Their recent work on combining aspects of paraconsistency and probability based on the logics of formal inconsistency [25] is particularly relevant to ENoRME. Both have collaborated before with van der Torre's group at U Luxembourg.

Broersen, who has collaborated with van der Torre in the past, adds further expertise on the combination of logics, e.g., the Beliefs-Obligations-Intentions-Desires architecture [22], and deontic epistemic logic [21]. He has also a deep interest in conceptual problems that arise from connecting a symbolic reasoning layer with a non-symbolic ML layer in ethical governors. This interest is shared by other consortium members, including Rojas, DLR and Latentine GmbH.

Walsh is a leading and an extraordinarily visible international ambassador of "Trustworthy AI". He is, among others, in close interaction with van der Torre and Benzmüller, on the highly controversial topic of Lethal Autonomous Weapons Systems (LAWS).

The research activities described above provide foundation stones for ENoRME. They will be consolidated, extended, and integrated with the work of ENoRME with the objective of developing a practical infrastructure for normative reasoning with exemplary deployment and experimentation. ENoRME will integrate the expertise of the consortium members to produce results (i) on the methodological level of explicit ethical governing architectures, (ii) that widen the scope of logics and logic combinations required in non-trivial normative reasoning, and (iii) demonstrate applicability in explorative applications such as autonomous driving, banking, finance, social and cognitive robotics, marketing. To ensure scalability and availability for future applications beyond ENoRME, a cloud-based, on-demand, URW will be developed and deployed (in addition to more traditional solutions based on single computer and cluster installations). These activities will ensure that ENoRME produces practical tools that can be deployed in applications, and at the same time rest on proper conceptual and theoretical foundations.

4.1 Work Streams

The ENoRME activities will be organised in the following parallel work streams that will be refined into work packages with individual schedules, milestones, and deliverables. Due to space restrictions we cannot discuss the details of the work packages here, but have to postpone this to the second stage of the evaluation process. There exist some dependencies between the work streams, which we are aware of and which are reflected in the order of presentation of the work streams. These dependencies will be used to guide work scheduling. Nevertheless, nearly all streams have components that can start right away, and all researchers will be able to contribute from the beginning. The lead person of each work stream is highlighted in italics, and overall project coordination will be under *Benzmüller* and Fuenmayor.

WS-1: Logics and Logic Combinations. We will address challenges of knowledge representation and reasoning with flexible combinations of highly expressive non-classical logics. We will carry out foundational research on the automation of reasoning with inconsistent and incomplete information, and the automation of logic combinations (e.g., deontic, epistemic, probabilistic, paraconsistent), supported by higher-order reasoning frameworks. **Responsible:** Benzmüller, Bueno-Soler, *Carnielli*, Parent; **Contributors:** Broersen, Libal, van der Torre.

WS-2: Standardisation, Systems Integration and Benchmarks. We will standardise machine and human readable formats for expressing problems and solutions in many different logics and logic combinations. Sutcliffe is a leading expert in this regard with his successful TPTP project. We will develop benchmarking tools for comparative assessment of reasoning tools. This work will generalise, adapt and improve existing TPTP standards, tooling, and infrastructure [90]. **Responsible:** Fuenmayor, Steen, *Sutcliffe*; **Contributors:** DLR, Latentine GmbH, Libal, Schulz.

WS-3: Formalisation of Ethical and Legal Theories. We will model and encode selected legal and ethical theories using highly expressive logic combinations, thus leveraging the human-auditable, explicit format that is also amenable to normative reasoning and verification by automated means. **Responsible:** Benzmüller, Dennis, Fuenmayor, *Parent*, Slavkovic; **Contributors:** Broersen, DLR, KPMG Lighthouse Luxembourg, Pease, van der Torre, Wyner

WS-4: Cloud-based Universal Reasoning Workbench. We will develop an accessible, cloud-based platform supporting different reasoning paradigms and different logic, and logic combinations. We will use this platform to address the needs of SMEs and educational institutions, to experiment with AI at scale with little risk and low costs. Further R&D will be carried out in the areas of automated and interactive theorem proving and human-machine interfaces, with the aim of designing and implementing a human-centric, integrated user interface enabling effective user interactions with disparate AI systems. We will collaborate with the CLAIRE initiative [32] and plan to make our URW available within Europe's emerging "CERN for AI" [87]. **Responsible:** Fuenmayor, *Steen*, Sutcliffe; **Contributors:** wizAI solutions GmbH, Libal, Pease.

WS-5: Integration with Defeasible, Commonsense Reasoning and Cognitive Systems. We will model aspects of human cognition with the aim of creating systems able to simulate human rational thought processes. R&D activities will involve the fields of NLP, KR, automated reasoning, and ML. Since humans naturally reason in the presence of incomplete and inconsistent knowledge, the integration of results from **WS-1** is relevant here. **Responsible:** Bueno-Soler, Carnielli, Martinez, *Schon*; **Contributors:** DLR, LuxAI S.A.

WS-6: Integration with Data-Driven, Machine Learning Methods. We will address the integration of explicit, top-down modelling of ethical and legal constraints with bottom-up, data-driven learning approaches. This involves R&D in methods for extracting high-level symbolic representations from non-symbolic models. This work stream thus addresses a general "hot topic" in AI, but from the particular angle of reasoning with normative knowledge. **Responsible:** Benzmüller, Göhring, *Rojas*; **Contributors:** Broersen, Latentine GmbH, Schulz.

WS-7: Methodologies and Architectures for Ethical Governors. We will study the needs and requirements for ethical governors in selected application areas, and develop corresponding governing architectures. **Responsible:** Benzmüller, Dennis, Martinez, *Slavkovic*; **Contributors:** Broersen, DLR, KPMG Lighthouse Luxembourg, Latentine GmbH, LuxAI S.A., Fisher, Liao, van der Torre, Walsh, wizAI solutions GmbH, Wyner.

WS-8: Interaction and Communication with Ethical Governors. We will develop modes of human-centered interaction with ethical governors. **Responsible:** *Martinez*, Schon, Slavkovic; **Contributors:** Liao, LuxAI S.A., KPMG Lighthouse Luxembourg, van der Torre, Walsh, wizAI solutions GmbH, Wyner.

WS-9: Implementation, Experimentation and Demonstration. We will develop exemplary implementations of our technology in (at least) three ethically-critical AI areas involving the legal and ethical governing of (i) autonomous cars, (ii) algorithms in banking and finance, and (iii) social robots. Relevant bridge technologies required for the robust and secure implementation and assessment of our approach will be investigated. This work stream additionally includes experiments with simulated agent societies to study their behavior when constrained by legal or ethical theories. **Responsible:** Benzmüller, Dennis, Göhring, Martinez, Parent, Schon, *Steen*; **Contributors:** DLR, KPMG Lighthouse Luxembourg, Latentine GmbH, LuxAI S.A., Pease, Schulz, van der Torre, wizAI solutions GmbH.

WS-10: Verification of AI Systems. We will conduct studies towards establishing verifiably correct behaviour of IAS. In addition to legal and ethical governance, we address questions regarding best practices for secure implementation and deployment of AI systems. In particular, formal verification will realistically be possible for some critical AI subsystems. This work stream will conduct studies (in cooperation with local collaborators Prof. Roth and Prof. Margraf, FU Berlin) to assess these challenges. **Responsible:** Benzmüller, *Dennis*, Parent, Rojas, Steen; **Contributors:** DLR, Libal, Fisher.

WS-11: Education, Dissemination, Events and Media (Public AI Lab). This work involves the collaboration with projects/doctoral schools (e.g. in Berlin and Luxembourg) as well as the organization of lectures, workshops, conferences and e-Learning courses to train a next generation of professionals in the topics of ENoRME by leveraging the cloud-infrastructure developed in **WS-4** and the partnerships with the home institutions of the participants. We will closely interact with Gesellschaft für Informatik e.V. (science communication), Romeike (CS didactics, e.g., for interaction with schools in Berlin) and Walsh (international media presence). **Responsible:** *Benzmüller*, all participants; **Contributors:** Gesellschaft für Informatik e.V., Romeike, Walsh.

5 Sustainability of Collaboration

The existing collaborations between FU Berlin, U Luxembourg, and U Miami have already been mentioned. These collaborations will be further strengthened and expanded in various ways, including further joint research projects, joint teaching, joint PhD supervision. The cloud-based URW that will emerge from ENoRME, in combination with the related educational material (incl. e-Learning formats), is thereby adopting a key role as an enabling technology for practical AI research that is not possible otherwise. With this much needed technological contribution, ENoRME is reaching out to both theoreticians and practitioners, in AI, computer science, robotics, and philosophy, as reflected in the composition of the ENoRME consortium. ENoRME thus provides bridge technology to leverage relevant interdisciplinary research towards Trustworthy AI and beyond. From the start ENoRME will actively reach out to interested users. Headed by Sutcliffe, we plan to widen the scope of the yearly CASC world championship [94] for classical ATP to cover non-classical logics as addressed in ENoRME. This will significantly speed-up the development of robust, high-performance tools, as required in ENoRME and in IASs, and in further fields such as formal argumentation [49], philosophy, and computational metaphysics [62]. As such, ENoRME will develop the bridge technology that is needed to connect theoreticians and practitioners in AI and beyond. The core technological developments of ENoRME are intended to become part of the emerging "CERN for AI" [87], to guarantee their availability beyond the scope of ENoRME.

6 Communication Activities and Knowledge Transfer

The three main pillars of ENoRME's public communication strategy provide a wide spectrum of different consumption channels for scientists, working professionals, interested amateurs, and students. They are:

Extensive Web Presence. A modern web presence will give detailed information about the goals, progress, current results, and events of ENoRME. Important findings and statements will be presented as suitable press releases and regular e-mail newsletters that are available to the public by subscription. We will collaborate closely with the participating institutions and the press for increased outreach. One of the main features of ENoRME's web presence will be the provision of the cloud-based URW that can be used freely to run experiments in AI ethics. Examples, demonstrations, and teaching material will be prepared. Similarly, suitable material will be provided for school and university teaching. E-Learning formats will also be targeted, in particular a Massive Open Online Course (MOOC) will be developed with the support, among others, of the Gesellschaft für Informatik e.V.. Visibility in social media is of strategic importance to our publicity objectives. We will have a dedicated community manager to help us develop and maintain a consistent multi-channel social media outreach, involving general social media channels (Facebook, Twitter, etc.), professional social media (ResearchGate, LinkedIn, etc.) and technology news sites.

Public AI Lab. The public AI lab constitutes the centrepiece of ENoRME's public communication efforts. It will be a curated permanent exhibition at FU Berlin that is freely accessible to the public. The lab will present static displays (e.g., posters and pictures) about the project and its research, and interactive hands-on exhibitions that allow visitors to explore the research questions of ENoRME. At regular intervals the ENoRME team will offer guided tours of the AI lab (e.g., for schools in Berlin[10]), professionally supervise advanced group experiments, and offer consultancy (e.g., to SMEs and startups). The AI lab will also host relevant university teaching associated with ENoRME (see further below). For the successful implementation of the public AI lab, ENoRME will benefit from existing local expertise and

[10] A Guided Tour through the Dahlem Center for Machine Learning and Robotics (DCMLR) for children from the Kronach-Grundschule in Berlin-Lichterfelde was e.g. organised on June 6, 2019; at the DCMLR we have organised various similar events before.

infrastructure, e.g., in MATH+ (formerly Matheon). Romeike, an expert in the didactics in computer science, and Krupka, a science communication expert and member of the executive board of the Gesellschaft für Informatik e.V., will support ENoRME regarding media outreach, the organization of a MOOC, and the installation of the AI Lab.

Regular Events. The core portfolio of ENoRME events includes both regular scientific meetings as well as non-scientific community events. Regular scientific project workshops, including a festive public kick-off event in Berlin, will be hosted at FU Berlin and at participating institutions. Additionally, we plan to organize one edition of an international high-visibility conference (such as CADE, IJCAR, ECAI, or IJCAI) at FU Berlin, similar to e.g., [16] and [11]. For each of the scientific events, dedicated proceedings will be published in an open-access venue. For low-threshold introductions to ethical AI research and the education of prospective young researchers, ENoRME will contribute to university teaching at the Master's level. Drawing from previous successes and experience with innovative lecture courses and teaching events [45], we plan to further develop our existing courses on Computational Metaphysics and Universal Logical Reasoning by exploiting the URW. Similar courses will also be offered at the home institutions of our visiting researchers and published (in a suitably adapted format) as online e-Learning resources that are freely available. ENoRME will present itself (and its AI lab) at local events, such as the annual *Lange Nacht der Wissenschaften* and further independently organized open days (*Tage der offenen Tür*, see e.g., [44]). Non-scientific meetings will include networking events for industry cooperation partners, public panel debates, and train-the-trainer sessions for AI professionals. A particular emphasis will be put on knowledge transfer with German industry and startups, through initiatives like the SME 4.0 Centres of Excellence and the European Digital SME Alliance. For the integration of topics relevant to society as a whole, we plan to make use of the well-established public lecture series (*Ringvorlesung*) of FU Berlin [46], and to provide a dedicated instance of that series related to different aspects of ENoRME – different actors from politics, industry, non-government organizations, and educational institutions will be invited to contribute. All events will be announced online (ENoRME's web page and social media), and advertised using multiple channels (local news, print media, newsletter announcements, etc.).

The ultimate goal of the project's communication strategy is to give insights into ENoRME developments, results, and applications at different levels of detail and, at the same time, promote Berlin/Germany, Luxembourg and Europe as a vivid research and development site for cutting-edge AI research into critical reflection and balanced AI technology that simultaneously invests in innovation and risk prevention.

ENoRME results will be continuously presented at internationally recognized conferences, and archived in peer reviewed journals. A soft goal is to prepare a textbook about machine ethics, automated reasoning, and AI, which can be used as self-contained teaching material.

7 Estimated Budget

Estimated Total Budget	€4,990,000
Personnel (see attached Excel sheet for further details) *Includes: (i) participating researchers, (ii) associated supporting young researchers (6 months per participating researcher), (iii) software architect/lead of cloud-based URW, (iv) student assistants (HiWi) for URW development (web/backend, systems integration, etc.), (v) student assistants (HiWi) for project coordination support (conferences, workshops, e-learning, PR/community management, etc.), (vi) community manager & coordination of public AI lab*	€3,900,000
Housing Costs *Max. of €1000 per month per incoming participating researcher*	€250,000
Hardware, Software, Cloud-Infrastructure, MOOC Production, Public AI Lab	€120,000
Travel Costs *Includes: (i) incoming travel of participating researchers, (ii) conference/workshop travel of project members and supervised students, (iii) incoming travel and accommodation of further collaborators of the project*	€250,000
Workshop and Conference Organisation	€50,000
Office Space Rental	€350,000
Website, Media, Publication (incl. Open Access), further Dissemination	€40,000
Consumables, Printing, Office Equipment, Office Supply	€30,000

References

[1] M. Anderson and S. L. Anderson. Toward ensuring ethical behavior from autonomous systems: a case-supported principle-based paradigm. *Industrial Robot*, 42(4):324–331, 2015.

[2] R. Arkin, P. Ulam, and B. Duncan. An ethical governor for constraining lethal action in an autonomous system. Technical Report GIT-GVU-09-02, Mobile Robot Laboratory, College of Computing, Georgia Institute of Technology, 2009.

[3] R. Arkin, P. Ulam, and A. Wagner. Moral decision making in autonomous systems: Enforcement, moral emotions, dignity, trust, and deception. *Proceedings of the IEEE*, 100(3):571–589, 2012.

[4] M. Bender, B. Pelzer, and C. Schon. System description: E-KRHyper 1.4 – extensions for unique names and description logic. In M. P. Bonacina, editor, *Automated Deduction – CADE-24 – 24th International Conference on Automated Deduction, Lake Placid, NY, USA, June 9-14, 2013. Proceedings*, volume 7898 of *Lecture Notes in Computer Science*, pages 126–134. Springer, 2013.

[5] C. Benzmüller. Universal (meta-)logical reasoning: Recent successes. *Science of Computer Programming*, 172:48–62, 2019. http://dx.doi.org/10.1016/j.scico.2018.10.008.

[6] C. Benzmüller. Universal (meta-)logical reasoning: The wise men puzzle (Isabelle/HOL dataset). *Data in Brief*, 24(103774), 2019. Note: data publication http://dx.doi.org/10.1016/j.dib.2019.103823.

[7] C. Benzmüller and P. Andrews. Church's type theory. In E. N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, summer 2019 edition, 2019.

[8] C. Benzmüller, A. Farjami, P. Meder, and X. Parent. I/O logic in HOL. *J. of Applied Logics – IfCoLoG J. of Logics and their Applications*, 2019. To appear, URL (preprint): https://www.researchgate.net/publication/332786587_IO_Logic_in_HOL.

[9] C. Benzmüller, A. Farjami, and X. Parent. A dyadic deontic logic in HOL. In J. Broersen, C. Condoravdi, S. Nair, and G. Pigozzi, editors, *Deontic Logic and Normative Systems — 14th International Conference, DEON 2018, Utrecht, The Netherlands, 3-6 July, 2018*, volume 9706, pages 33–50. College Publications, 2018. John-Jules Meyer Best Paper Award.

[10] C. Benzmüller, A. Farjami, and X. Parent. Åqvist's dyadic deontic logic E in HOL. *J. of Applied Logics – IfCoLoG J. of Logics and their Applications*, 2019. To appear, URL (preprint): https://www.researchgate.net/publication/332786724_Aqvist's_Dyadic_Deontic_Logic_E_in_HOL.

[11] C. Benzmüller, X. Parent, and F. Ricca. RuleML+RR (Rules and Reasoning Symposium) 2018 Report. *AI Magazine*, 2019. To appear, preprint: http://doi.org/10.13140/RG.2.2.18326.19529.

[12] C. Benzmüller, X. Parent, and L. van der Torre. A deontic logic reasoning infrastructure. In F. Manea, R. G. Miller, and D. Nowotka, editors, *14th Conference on Computability in Europe, CiE 2018, Kiel, Germany, July 30-August, 2018, Proceedings*, volume 10936 of *Lecture Notes in Computer Science*, pages 60–69. Springer, 2018.

[13] C. Benzmüller, X. Parent, and L. van der Torre. Designing normative theories of ethical reasoning: Formal framework, methodology, and tool support. *preprint of (submitted) article*, 2019. Url (preprint): http://doi.org/10.13140/RG.2.2.10502.42561.

[14] C. Benzmüller, F. Rabe, and G. Sutcliffe. THF0 – the core of the TPTP language for classical higher-order logic. In A. Armando, P. Baumgartner, and G. Dowek, editors, *Automated Reasoning, 4th International Joint Conference, IJCAR 2008, Sydney, Australia, August 12-15, 2008, Proceedings*, volume 5195 of *Lecture Notes in Computer Science*, pages 491–506. Springer, 2008. Url (preprint): http://christoph-benzmueller.de/papers/C25.pdf.

[15] C. Benzmüller and D. S. Scott. Automating free logic in HOL, with an experimental application in category theory. *J. of Automated Reasoning*, 2019. To appear, online version: http://doi.org/10.1007/s10817-018-09507-7.

[16] C. Benzmüller, A. Steen, and M. Wisniewski. News — 25th International Conference on Automated Deduction (CADE-25). *Künstliche Intelligenz 29(4):451-452; conference report*, 2015.

[17] J.-Y. Béziau, W. A. Carnielli, and D. M. Gabbay. *Handbook of paraconsistency*. College publications, 2007.

[18] E. P. Bjørgen, S. Madsen, T. S. Bjørknes, F. V. Heimsæter, R. Håvik, M. Linderud, P.-N. Longberg, L. A. Dennis, and M. Slavkovik. Cake, death, and trolleys: dilemmas as benchmarks of ethical decision-making. In *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*, pages 23–29. ACM, 2018.

[19] P. Bremner, L. A. Dennis, M. Fisher, and A. F. T. Winfield. On proactive, transparent, and verifiable ethical reasoning for robots. *Proceedings of the IEEE*, 107(3):541–561, 2019.

[20] S. Bringsjord, N. Sundar G., B. F. Malle, and M. Scheutz. Contextual deontic cognitive event calculi for ethically correct robots. In *Proceedings of the International Symposium on Artificial Intelligence and Mathematics (ISAIM)*, 2018.

[21] J. Broersen. Deontic epistemic stit logic distinguishing modes of mens rea. *J. of Applied Logic*, 9(2):137 – 152, 2011. Special Issue: Selected and revised papers from the Ninth International Conference on Deontic Logic in Computer Science (DEON 2008).

[22] J. Broersen, M. Dastani, and L. van der Torre. Beliefs, obligations, intentions, and desires as components in an agent architecture. *Int. J. of Intelligent Systems*, 20(9):893–919, 2005.

[23] C. E. Brown, T. Gauthier, C. Kaliszyk, G. Sutcliffe, and J. Urban. GRUNGE: A grand unified ATP challenge. *CoRR*, abs/1903.02539, 2019. Accepted for CADE 2019 (preprint): http://arxiv.org/abs/1903.02539.

[24] J. Bueno-Soler. Two semantical approaches to paraconsistent modalities. *Logica Universalis*, 4(1):137–160, 2010.

[25] J. Bueno-Soler and W. Carnielli. Paraconsistent probabilities: Consistency, contradictions and Bayes' theorem. *Entropy*, 18(9):325, 2016.

[26] J. Bueno-Soler and W. Carnielli. Experimenting with consistency. In V. Markin and D. Zaitsev, editors, *The Logical Legacy of Nikolai Vasiliev and Modern Logic*, volume 387 of *Synthese Library (Studies in Epistemology, Logic, Methodology, and Philosophy of Science)*. Springer, Cham, 2017.

[27] J. Bueno-Soler and W. A. Carnielli. Possible-translations algebraization for paraconsistent logics. *Bulletin of the Section of Logic*, 34(2):77–92, 2005.

[28] H. D. Burkhard, D. Duhaut, M. Fujita, P. Lima, R. Murphy, and R. Rojas. The road to robocup 2050. *IEEE Robotics Automation Magazine*, 9(2):31–38, June 2002.

[29] W. Carnielli, M. Coniglio, D. Gabbay, P. Gouveia, and C. Sernadas. *Analysis and Synthesis of Logics – How to Cut and Paste Reasoning Systems*, volume 35 of *Applied Logic Series*. Springer Netherlands, 2008.

[30] W. Carnielli and M. E. Coniglio. *Paraconsistent Logic: Consistency, Contradiction and Negation*, volume 40 of *Logic, Epistemology, and the Unity of Science*. Springer International Publishing, 2016.

[31] W. Carnielli, M. E. Coniglio, and J. Marcos. Logics of formal inconsistency. In *Handbook of philosophical logic*, pages 1–93. Springer, 2007.

[32] CLAIRE. Confederation of Laboratories for Artificial Intelligence Research in Europe, 2018. https://claire-ai.org.

[33] E. V. Cuevas, D. Zaldivar, and R. Rojas. Kalman filter for vision tracking. 2005. https://refubium.fu-berlin.de/handle/fub188/19186.

[34] L. A. Dennis and M. Fisher. Practical challenges in explicit ethical machine reasoning. In *International Symposium on Artificial Intelligence and Mathematics, ISAIM 2018, Fort Lauderdale, Florida, USA, January 3-5, 2018.*, 2018.

[35] L. A. Dennis, M. Fisher, M. Slavkovik, and M. Webster. Formal verification of ethical choices in autonomous systems. *Robotics and Autonomous Systems*, 77:1–14, 2016.

[36] L. A. Dennis, M. Fisher, M. P. Webster, and R. H. Bordini. Model checking agent programming languages. *Automated software engineering*, 19(1):5–63, 2012.

[37] M. Deters, A. Reynolds, T. King, C. W. Barrett, and C. Tinelli. A tour of CVC4: How it works, and how to use it. In K. Claessen and V. Kuncak, editors, *Formal Methods in Computer-Aided Design, FMCAD 2014, Lausanne, Switzerland, October 21-24, 2014*, page 7. IEEE, 2014.

[38] V. Dignum. Responsible autonomy. In *Proceedings of IJCAI-17*, pages 4698–4704, 2017.

[39] V. Dignum, M. Baldoni, C. Baroglio, M. Caon, R. Chatila, L. A. Dennis, G. Génova, G. Haim, M. S. Kließ, M. López-Sánchez, R. Micalizio, J. Pavón, M. Slavkovik, M. Smakman, M. van Steenbergen, S. Tedeschi, L. van der Torre, S. Villata, and T. de Wildt. Ethics by design: Necessity or curse? In *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society, AIES 2018, New Orleans, LA, USA, February 02-03, 2018*, pages 60–66, 2018.

[40] V. Dignum (ed.). Special issue: Ethics and artificial intelligence. *Ethics and Information Technology*, 20(1), 2018.

[41] European Commission, Independent High-Level Expert Group on Artificial Intelligence. Ethics Guidelines for Thrustworthy AI. https://ec.europa.eu/futurium/en/ai-alliance-consultation, April 2019.

[42] M. Fisher, L. Dennis, and M. Webster. Verifying autonomous systems. *Communications of the ACM*, 56(9):84–93, 2013.

[43] L. Floridi. What the near future of artificial intelligence could be. *Philosophy & Technology*, 32(1):1–15, 2019.

[44] FU Berlin Press Release. Tag der Offenen Tür am 4. Juni 2019 für Journalistinnen und Journalisten am Dahlem Center for Machine Learning and Robotics, June 2019. https://www.mi.fu-berlin.de/inf/groups/ag-ki/Dates/Tag-der-offenen-Tuer.html.

[45] FU Berlin Press Release (Nr. 022/2016). Lehrpreis 2015 der Freien Universität Berlin: Logikausbildung 3.0, 2016. https://www.fu-berlin.de/presse/informationen/fup/2016/fup_16_022-lehrpreis-benzmueller-logik-gottesbeweis/index.html.

[46] FU Berlin Press Release (Nr. 066/2019). Der offene Hörsaal im Sommersemester, 2019. https://www.fu-berlin.de/presse/informationen/fup/2019/fup_19_066-offener-hoersaal-sommersemester-2019-ueberblick/index.html, see also https://www.fu-berlin.de/sites/offenerhoersaal.

[47] D. Fuenmayor and C. Benzmüller. A case study on computational hermeneutics: E. J. Lowe's modal ontological argument. *J. of Applied Logic - IfCoLoG J. of Logics and their Applications (special issue on Formal Approaches to the Ontological Argument)*, 5(7):1567–1603, 2018. To be published also as chapter in the book 'Beyond Faith and Rationality: Essays on Logic, Religion and Philosophy' printed in the Springer book series 'Sophia Studies in Cross-cultural Philosophy of Traditions and Cultures', preprint: http://christoph-benzmueller.de/papers/J38.pdf.

[48] D. Fuenmayor and C. Benzmüller. Formalisation and evaluation of Alan Gewirth's proof for the Principle of Generic Consistency in Isabelle/HOL. *Archive of Formal Proofs*, 2018. Formally verified data publication. https://www.isa-afp.org/entries/GewirthPGCProof.html.

[49] D. Fuenmayor and C. Benzmüller. Computational hermeneutics: An integrated approach for the logical analysis of natural-language arguments. In B. Liao, T. Agotnes, and Y. N. Wang, editors, *Postproceedings of CLAR-2018*, Logic in Asia series (Studia Logica Library). Springer Singapore, 2019. Url (preprint): http://doi.org/10.13140/RG.2.2.11784.26881, (accepted for publication).

[50] D. Fuenmayor and C. Benzmüller. Harnessing higher-order (meta-)logic to represent and reason with complex ethical theories. In *PRICAI 2019: Trends in Artificial Intelligence*, Lecture Notes in Artificial Intelligence. Springer International Publishing, 2019. Preprint http://arxiv.org/abs/1903.09818.

[51] U. Furbach, S. Hölldobler, M. Ragni, and C. Schon. Workshop: Bridging the gap: Is logic and automated reasoning a foundation for human reasoning? In G. Gunzelmann, A. Howes, T. Tenbrink, and E. J. Davelaar, editors, *Proceedings of the 39th Annual Meeting of the Cognitive Science Society, CogSci 2017, London, UK, 16-29 July 2017*. cognitivesciencesociety.org, 2017.

[52] U. Furbach, T. Krämer, and C. Schon. Names are not just sound and smoke: Word embeddings for axiom selection. In *Proceedings of CADE-27 – The 27th International Conference on Automated Deduction*, 2019. To appear, preprint: https://www.bibsonomy.org/documents/f137cd29d14b51ec7a825a7a9832ac6d/cschon/cade.pdf.

[53] U. Furbach, B. Pelzer, and C. Schon. Automated reasoning in the wild. In A. P. Felty and A. Middeldorp, editors, *Automated Deduction – CADE-25 – 25th International Conference on Automated Deduction, Berlin, Germany, August 1-7, 2015, Proceedings*, volume 9195 of *Lecture Notes in Computer Science*, pages 55–72. Springer, 2015.

[54] U. Furbach and C. Schon. Deontic logic for human reasoning. In T. Eiter, H. Strass, M. Truszczynski, and S. Woltran, editors, *Advances in Knowledge Representation, Logic Programming, and Abstract Argumentation*, volume 9060 of *Lecture Notes in Computer Science*, pages 63–80. Springer, 2015.

[55] U. Furbach and C. Schon. Commonsense reasoning meets theorem proving. In M. Klusch, R. Unland, O. Shehory, A. Pokahr, and S. Ahrndt, editors, *Multiagent System Technologies - 14th German Conference, MATES 2016, Klagenfurt, Österreich, September 27-30, 2016. Proceedings*, volume 9872 of *Lecture Notes in Computer Science*, pages 3–17. Springer, 2016.

[56] D. Gabbay, J. Horty, X. Parent, R. van der Meyden, and L. van der Torre. *Handbook of deontic logic and normative systems*. College Publication, 2013.

[57] F. R. Gallo, G. I. Simari, M. V. Martinez, M. A. Falappa, and N. A. Santos. Reasoning about sentiment and knowledge diffusion in social networks. *IEEE Internet Computing*, 21(6):8–17, 2017.

[58] T. Gleißner, A. Steen, and C. Benzmüller. Theorem provers for every normal modal logic. In T. Eiter and D. Sands, editors, *LPAR-21. 21st International Conference on Logic for Programming, Artificial Intelligence and Reasoning*, volume 46 of *EPiC Series in Computing*, pages 14–30, Maun, Botswana, 2017. EasyChair.

[59] D. Göhring, D. Latotzky, M. Wang, and R. Rojas. Semi-autonomous car control using brain computer interfaces. In *Intelligent Autonomous Systems 12*, pages 393–408. Springer, 2013.

[60] S. Guadarrama, L. Riano, D. Golland, D. Göhring, Y. Jia, D. Klein, P. Abbeel, T. Darrell, et al. Grounding spatial relations for human-robot interaction. In *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1640–1647. IEEE, 2013.

[61] J. Hoffman, M. Spranger, D. Göhring, and M. Jüngel. Making use of what you don't see: Negative information in Markov localization. In *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 2947–2952. IEEE, 2005.

[62] D. Kirchner, C. Benzmüller, and E. N. Zalta. Computer science and metaphysics: A cross-fertilization. *Open Philosophy (Special Issue – Computer Modeling in Philosophy)*, 2019. To appear, preprint: http://doi.org/10.13140/RG.2.2.25229.18403.

[63] E. Kontopoulos, N. Bassiliades, G. Governatori, and G. Antoniou. A modal defeasible reasoner of deontic logic for the semantic web. *Int. J. Semantic Web Inf. Syst.*, 7(1):18–43, 2011.

[64] L. Kovács and A. Voronkov. First-Order Theorem Proving and Vampire. In N. Sharygina and H. Veith, editors, *Computer Aided Verification – 25th International Conference, CAV 2013, Saint Petersburg, Russia, July 13-19, 2013. Proceedings*, volume 8044 of *Lecture Notes in Computer Science*, pages 1–35. Springer, 2013.

[65] P. Kügler, P. Kestel, C. Schon, M. Marian, B. Schleich, S. Staab, and S. Wartzack. Ontology-based approach for the use of intentional forgetting in product development. In D. Marjanovic, M. Storga, N. Pavkovic, N. Bojcetic, and S. Skec, editors, *DESIGN 2018 – 15th International Design Conference (Dubrovnik, 05/21/18 - 05/24/18)*, 2018.

[66] J. Lang, M. Slavkovik, and S. Vesic. Agenda separability in judgment aggregation. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, AAAI'16, pages 1016–1022. AAAI Press, 2016.

[67] D. D. Lee, A. Steen, and T. Walsh, editors. *GCAI-2018, 4th Global Conference on Artificial Intelligence, Luxembourg, September 18-21, 2018*, volume 55 of *EPiC Series in Computing*. EasyChair, 2018.

[68] B. Liao, N. Oren, L. van der Torre, and S. Villata. Prioritized norms in formal argumentation. *J. of Logic and Computation*, 29(2):215–240, 2019.

[69] B. Liao, M. Slavkovik, and L. van der Torre. Building Jiminy Cricket: An architecture for moral agreements among stakeholders. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society, AIES 2019*, 2019.

[70] T. Lukasiewicz, M. V. Martinez, and G. I. Simari. Inconsistency handling in datalog+/-ontologies. In *Proceedings of the 20th European Conference on Artificial Intelligence (ECAI 2012)*, pages 558–563, 2012.

[71] B. F. Malle. Integrating robot ethics and machine morality: The study and design of moral competence in robots. *Ethics and Information Technology*, 18(4):243–256, 2016.

[72] M. V. Martinez. Knowledge engineering for intelligent decision support. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI 2017)*, pages 5131–5135, 2017.

[73] T. Nipkow, L. C. Paulson, and M. Wenzel. *Isabelle/HOL: A Proof Assistant for Higher-Order Logic*. Number 2283 in Lecture Notes in Computer Science. Springer, 2002.

[74] OECD. *Recommendation of the Council on Artificial Intelligence*. OECD/LEGAL/0449, https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449, 2019.

[75] J. Otten. Non-clausal connection calculi for non-classical logics. In R. A. Schmidt and C. Nalon, editors, *Automated Reasoning with Analytic Tableaux and Related Methods - 26th International Conference, TABLEAUX 2017, Brasília, Brazil, September 25-28, 2017, Proceedings*, volume 10501 of *Lecture Notes in Computer Science*, pages 209–227. Springer, 2017.

[76] X. Parent. Maximality vs. optimality in dyadic deontic logic. *J. Philosophical Logic*, 43(6):1101–1128, 2014.

[77] X. Parent. Completeness of Åqvist's systems E and F. *Review of Symbolic Logic*, 8(1):164–177, 2015.

[78] L. M. Pereira and A. Saptawijaya. *Programming Machine Ethics*, volume 26 of *Studies in Applied Philosophy, Epistemology and Rational Ethics*. Springer, 2016.

[79] L. M. Pereira and A. Saptawijaya. Counterfactuals, logic programming and agent morality. In R. Urbaniak and G. Payette, editors, *Applied Formal/Mathematical Philosophy*, Logic, Argumentation & Reasoning. Springer, 2017.

[80] B. Reese. *The Fourth Age: Smart Robots, Conscious Computers, and the Future of Humanity*. Atria Books, 2018.

[81] R. Rojas. *Neural Networks - A Systematic Introduction*. Springer, 1996.

[82] A. Saptawijaya and L. M. Pereira. Logic programming for modeling morality. *Logic J. of the IGPL*, 24(4):510–525, 2016.

[83] M. Scheutz. The case for explicit ethical agents. *AI Magazine*, 38(4):57–64, 2017.

[84] R. A. Schmidt and U. Hustadt. First-order resolution methods for modal logics. In *Programming Logics*, volume 7797 of *Lecture Notes in Computer Science*, pages 345–391. Springer, 2013.

[85] C. Schon, S. Siebert, and F. Stolzenburg. The CoRg project – cognitive reasoning. *KI*, 33(3), 2019. To appear, preprint: https://www.bibsonomy.org/documents/2e08f404b02ea69db9511b7ba86e3dc6/cschon/CoRg.pdf.

[86] S. Schulz. System description: E 1.8. In K. L. McMillan, A. Middeldorp, and A. Voronkov, editors, *Logic for Programming, Artificial Intelligence, and Reasoning - 19th International Conference, LPAR-19, Stellenbosch, South Africa, December 14-19, 2013. Proceedings*, volume 8312 of *Lecture Notes in Computer Science*, pages 735–743. Springer, 2013.

[87] P. Slussalek. Position paper on “CERN for AI”. Technical report, OECD, Oct. 2017. https://www.oecd-forum.org/users/71431-philipp-slusallek/posts/28452-artificial-intelligence-and-digital-reality-do-we-need-a-cern-for-ai.

[88] A. Steen. *Extensional Paramodulation for Higher-Order Logic and its Effective Implementation Leo-III*, volume 345 of *DISKI – Dissertations in Artificial Intelligence*. Akademische Verlagsgesellschaft AKA GmbH, Berlin, 9 2018. Dissertation, Freie Universität Berlin, Germany.

[89] A. Steen and C. Benzmüller. The higher-order prover Leo-III. In D. Galmiche, S. Schulz, and R. Sebastiani, editors, *Automated Reasoning. IJCAR 2018*, volume 10900 of *Lecture Notes in Computer Science*, pages 108–116. Springer, Cham, 2018.

[90] G. Sutcliffe. The TPTP problem library and associated infrastructure. *J. of Automated Reasoning*, 43(4):337, 2009.

[91] G. Sutcliffe. The 9th IJCAR automated theorem proving system competition - CASC-J9. *AI Communications*, 31(6):495–507, 2018.

[92] G. Sutcliffe and C. Benzmüller. Automated reasoning in higher-order logic using the TPTP THF infrastructure. *J. of Formalized Reasoning*, 3(1):1–27, 2010. Url (preprint): http://christoph-benzmueller.de/papers/J22.pdf.

[93] G. Sutcliffe, C. Benzmüller, C. Brown, and F. Theiss. Progress in the development of automated theorem proving for higher-order logic. In R. Schmidt, editor, *Automated Deduction - CADE-22, 22nd International Conference on Automated Deduction, Montreal, Canada, August 2-7, 2009. Proceedings*, volume 5663 of *Lecture Notes in Computer Science*, pages 116–130. Springer, 2009.

[94] G. Sutcliffe and C. Suttner. The state of CASC. *AI Communications*, 19(1):35–48, 2006.

[95] M. Tegmark. *Life 3.0: Being Human in the Age of Artificial Intelligence*. Knopf, 2017.

[96] The Federal Government of Germany. Artificial intelligence strategy. https://www.ki-strategie-deutschland.de, November 2018.

[97] I. J. Timm, S. Staab, M. Siebers, C. Schon, U. Schmid, K. Sauerwald, L. Reuter, M. Ragni, C. Niederée, H. Maus, G. Kern-Isberner, C. Jilek, P. Friemann, T. Eiter, A. Dengel, H. Dames, T. Bock, J. O. Berndt, and C. Beierle. Intentional forgetting in artificial intelligence systems: Perspectives and challenges. In F. Trollmann and A. Turhan, editors, *KI 2018: Advances in Artificial Intelligence - 41st German Conference on AI, Berlin, Germany, September 24-28, 2018, Proceedings*, volume 11117 of *Lecture Notes in Computer Science*, pages 357–365. Springer, 2018.

[98] P. Vukmirovic, J. C. Blanchette, S. Cruanes, and S. Schulz. Extending a brainiac prover to lambda-free higher-order logic. In T. Vojnar and L. Zhang, editors, *Tools and Algorithms for the Construction and Analysis of Systems - 25th International Conference, TACAS 2019, Held as Part of the European Joint Conferences on Theory and Practice of Software, ETAPS 2019, Prague, Czech Republic, April 6-11, 2019, Proceedings, Part I*, volume 11427 of *Lecture Notes in Computer Science*, pages 192–210. Springer, 2019.

[99] W. Wallach, C. Allen, and I. Smit. Machine morality: bottom-up and top-down approaches for modelling human moral faculties. *AI & Society*, 22(4):565–582, 2008.

[100] A. F. T. Winfield and M. Jirotka. Ethical governance is essential to building trust in robotics and artificial intelligence systems. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 376(2133):20180085, 2018.

Questions and Answers about ENoRME

Q_1 **Can machines really be ethical?** "Real" ethical behaviour in intelligent autonomous systems is presumably impossible without consciousness, a physical body, the feel of pain and joy, etc. Such a debate is, however, independent of the objectives of the ENoRME project, since we are not claiming that IASs can or even should have their "own ethics". Instead ENoRME is claiming that IASs should be governed by "our" moral norms and legal contraints.

Q_2 **Are there existing reasoning frameworks already addressing the challenges of ENoRME?** Existing tools are lacking in many ways. Among their shortcomings we count: restricted expressivity (e.g. restricted to propositional logics only, no support for non-classical logics); restriction to reasoning within a single logic (i.e. no support for flexible logic combinations); difficult integration within a larger framework (e.g. tooling and documentation not easily available, infrequent updates/bugfixes, no standardised input/output syntax); no possibilities for user-interaction; poor automation and poor explainability (e.g. lacking generation of readable proof objects). ENoRME aims at addressing these shortcomings.

Q_3 **What recent developments suggest that the goals of ENoRME can actually be tackled now?** Explicit normative reasoning is by no means trivial. In particular, the automation of quantified deontic logics is challenging. However, some significant progress has recently been achieved, among others, based on the universal logical reasoning approach [5], which fruitfully connects normative reasoning with state-of-the-art automated theorem proving technology. For initiating much needed further practical progress a cloud-based infrastructure and benchmarking means as envisioned in ENoRME will be essential. Competition and exemplary deployment will stimulate practical progress.

Q_4 **Why are the goals of ENoRME time critical? Could it be addressed also in ten years from now?** Intelligent and autonomous AI technology is developed at a high pace. Accompanying measures need to be developed early on and the defintion of Trustworthy AI should not be left to industry alone. The need for investment in this particular direction has been acknowledged by various leading researchers and institutions worldwide. Moreover, ENoRME contributes innovative, new technology that has numerous topical applications beyond this particular focus.

Q_5 **How good will the degree of proof automation be in the URW? Is it realistic to assume that ambitious normative reasoning can be automated to degree as required?** Prior work has shown encouraging results regarding proof automation. In particular, the granularity level as found in human normative reasoning can already be matched; in prior work we have revealed flaws and issues in refereed articles with our automated tools that have escaped the critical human eye.

Q_6 **How can non-symbolic and symbolic reasoning approaches be combined, especially in situations where fast decisions need to be made? If hybrid approaches are necessary, where would non-symbolic and symbolic AI go together, where would they act in a redundant or complementary way together?** This a core research question to be addressed in **WS-6—WS-9**, and the starting point will be related prior work and the leading experience of the team working on these topics. ENoRME will approach the challenge from both angles, theory and practice.

Q_7 **How will the collaboration with the industrial partners be implemented?** They will be implemented on the basis of co-supervised student projects (BSc, MSc, PhD), regular meetings, and joint events in Berlin and Luxembourg.

Q_8 **How will the industry partners, and industrial state-of-the-art in general, benefit from ENoRME? How can the results be utilized?** One of the principal tenets of ENoRME is the knowledge transfer of our research findings through our public AI Lab (see section 6) and leveraging our cloud-based Universal Reasoning Workbench (URW). We also aim at cooperating with Mittelstand 4.0 competence centres and the European Digital SME Alliance.

Q_9 **Which new (digital) services and business models might originate from the ENoRME project?** ENoRME wants to develop means for establishing highest levels of trust in emerging AI technology. If successful, this will provide various business opportunities like consulting services for industries, validation and 'callibration' of *Trustworthy AI* systems, and training and certification programs for AI professionals.

Q_{10} **How will it be ensured that the practical results and generated expertise of ENoRME will be available to German SME's?** Our cloud-based URW will be available beyond ENoRME. As deliverables we will produce training material (including a MOOC), guidelines and (reports on) exemplary case studies that will enable and inform later applications in SME's, industry and also academia.

Personnel Costs (Estimation)

Months per Pe	3	Level	Mean Salary																	
		W1	€5,819	DFG W1																
		W2	€7,861	DFG W2																
		W3	€11,677	DFG W3																
		VR	€1,750	see BMBF call																
		HiWi	€1,114	Stud-HW (80h per month)																
		Architect	€6,286	Lead software architect of the cloud platform - Level E13,3																
		PR	€4,992	Community manager and coordination of the public AI Lab - Level E9b,3																
										Core Period										
		Researcher	Level	20-Q1	20-Q2	20-Q3	20-Q4	21-Q1	21-Q2	21-Q3	21-Q4	22-Q1	22-Q2	22-Q3	22-Q4	#Periods	Cost			
	1	Christoph Benzmüller	W3	X	X	X	X	X	X	X	X	X	X	X	X	36	€420,363			
	2	Raúl Rojas	W3				X	X	X	X	X	X	X	X	X	27	€315,272		Housing	
	3	Daniel Göhring	W2					X	X	X	X	X	X	X	X	24	€188,664		allowance	1000
	4	Claudia Schon	W2						X	X	X	X	X	X	X	21	€165,081			21
	5	Xavier Parent	W2	X	X	X	X	X	X	X	X	X	X	X	X	36	€282,996			36
	6	Alexander Steen	W1	X	X	X	X	X	X	X	X	X	X	X	X	36	€209,498			36
	7	Marija Slavkovik	W2							X	X	X	X	X	X	18	€141,498			18
	8	Geoff Sutcliffe	W3			X	X	X	X	X	X	X	X			24	€280,242			24
	9	Louise Dennis	W2							X	X	X	X	X	X	18	€141,498			18
	10	Walter Carnielli	W3					X	X	X	X	X	X	X	X	24	€280,242			24
Research	11	Juliana Bueno-Soler	W2					X	X	X	X	X	X	X	X	24	€188,664			24
Team	12	Maria Vanina Martinez	W2			X	X	X	X	X	X	X	X			24	€188,664			24
	13	Visiting Researcher 1	VR							X	X					6	€10,500			6
	14	Visiting Researcher 2	VR							X	X					6	€10,500			6
	15	Visiting Researcher 3	VR							X	X					6	€10,500			6
	16	Visiting Researcher 4	VR							X	X					6	€10,500			6
	17	Visiting Researcher 5	VR							X	X					6	€10,500			6
	18	Visiting Researcher 6	VR							X	X					6	€10,500			6
	19	Visiting Researcher 7	VR							X	X					6	€10,500			6
	20	Visiting Researcher 8	VR							X	X					6	€10,500			6
	21	Visiting Researcher 9	VR							X	X					6	€10,500			6
	22	Visiting Researcher 10	VR							X	X					6	€10,500			6
	23	Visiting Researcher 11	VR							X	X					6	€10,500			6
	24	Visiting Researcher 12	VR							X	X					6	€10,500			6
Technology &	25	Cloud Platform Lead Architect	Architect	X	X	X	X	X	X	X	X	X	X	X	X	36	€226,284		Total	297000
Infrastructure	26	Web Development/UI Design	HiWi	X	X	X	X	X	X	X	X	X	X	X	X	36	€40,108			
Team	27	Web Development/UI Design	HiWi	X	X	X	X	X	X	X	X	X	X	X	X	36	€40,108			
	28	Backend/Cloud Development	HiWi	X	X	X	X	X	X	X	X	X	X	X	X	36	€40,108			
	29	Backend/Cloud Development	HiWi	X	X	X	X	X	X	X	X	X	X	X	X	36	€40,108		Personel	3,900,000
	30	AI Systems Integration	HiWi	X	X	X	X	X	X	X	X	X	X	X	X	36	€40,108		Housing	250,000
	31	AI Systems Integration	HiWi	X	X	X	X	X	X	X	X	X	X	X	X	36	€40,108		Cloud, infra, etc	120,000
	32	AI Systems Integration	HiWi	X	X	X	X	X	X	X	X	X	X	X	X	36	€40,108		Travel	250,000
	33	AI Systems Performance (HW)	HiWi	X	X	X	X	X	X	X	X	X	X	X	X	36	€40,108		Conferences, etc	50,000
	34	AI Systems Performance (SW)	HiWi	X	X	X	X	X	X	X	X	X	X	X	X	36	€40,108		Office rent	350,000
	35	IT infrastructure for AI Lab	HiWi	X	X	X	X	X	X	X	X	X	X	X	X	36	€40,108		Media, PR, etc.	40,000
Coordination	36	PR and coordination of AI Lab	PR	X	X	X	X	X	X	X	X	X	X	X	X	36	€179,700		Consumables	30,000
& Support	37	Visiting scientists support	HiWi	X	X	X	X	X	X	X	X	X	X	X	X	36	€40,108		Total	4,990,000
Team	38	Courses material, e-Learning	HiWi	X	X	X	X	X	X	X	X	X	X	X	X	36	€40,108			
	39	HR, internships, industry contact	HiWi	X	X	X	X	X	X	X	X	X	X	X	X	36	€40,108			
	40	Conferences, workshops, etc.	HiWi	X	X	X	X	X	X	X	X	X	X	X	X	36	€40,108			
																Total	€3,896,174			
Comments:		Scientists (W1, W2 & W3)	Role for scientists members of the Future Lab (subdivided in three levels according to seniority)																	
		Visiting Researchers (VR)	Each member scientist can bring one additional visiting researcher for up to 6 months (1750EUR per month)																	
		Student assistants (HiWi)	Students who work closely with and support their respective team coordinators/leads at the coordinating institution (part time job: 20h per week)																	
		Architect	Professional software architect with specialist AI knowledge. Tasks: Technical design, development and commisioning of the cloud-based platform integrating all of the FutureLab developments.																	
		PR	Professional in charge of Public Relations, Community management and the coordination of the public AI Lab (co-supervised with GI)																	

Cloud Computing Costs (Estimation)

All resources and prices are shown as an illustration (they are the ones provided by AWS by May/2019). Importantly: we strive towards a cloud-agnostic platform!

months per period	3
conversion USD/EUR	0.89

												core period								period after 01.2023		
env	**resource**	**comment**	**$USD/m**	**usage**	**# inst**	**20-Q1**	**20-Q2**	**20-Q3**	**20-Q4**	**21-Q1**	**21-Q2**	**21-Q3**	**21-Q4**	**22-Q1**	**22-Q2**	**22-Q3**	**22-Q4**	**months**	**initial cost**	**months**	**cost**	**total cost**
DEV	c5.large	8h per day	70	25%	8	X	X	X	X	X	X	X	X	X	X	X	X	36	$5,040	0	$0	$5,040
DEV	p3.2xlarge	GPU accel. ~72h per month ($3.06/hou	2200	10%	1	X	X	X	X	X	X	X	X	X	X	X	X	36	$7,920	0	$0	$7,920
DEV	f1.2xlarge	FPGA accel. ~72h per month ($1.65/hc	1200	10%	1	X	X	X	X	X	X	X	X	X	X	X	X	36	$4,320	0	$0	$4,320
TEST	p3.2xlarge	GPU accel. ~36h per month ($3.06/hou	2200	5%	1				X	X	X	X	X	X	X	X	X	27	$2,970	12	$1,320	$4,290
TEST	f1.2xlarge	FPGA accel. ~36h per month ($1.65/hc	1200	5%	1				X	X	X	X	X	X	X	X	X	27	$1,620	12	$720	$2,340
TEST	c5.xlarge	24/7	140	100%	4				X	X	X	X	X	X	X	X	X	27	$15,120	12	$6,720	$21,840
PROD	c5.xlarge	24/7	140	100%	8							X	X	X	X	X	X	18	$20,160	0	$0	$20,160
Others																					$0	$0
storage	~500 GB	distributed SSD-based storage among	60	100%	1	X	X	X	X	X	X	X	X	X	X	X	X	36	$2,160	12	$720	$2,880
I/O	~1 TB	distributed among all (~16) instances	50	100%	1	X	X	X	X	X	X	X	X	X	X	X	X	36	$1,800	12	$600	$2,400
support	Business	AWS Support (Business Plan)	100	100%	1	X	X	X	X	X	X	X	X	X	X	X	X	36	$3,600	0	$0	$3,600
																		Total $	$61,110		$10,080	$71,190
																		Total EUR	€54,388		€8,971	€63,359

Notes

DEV	Environment for daily development work. Very unstable, since artifacts are continuously (automatically) deployed and tested
TEST	Environment which developers and group researchers can use to run jobs, deploy proof-of-concepts, etc. More stable than DEV. Breaking changes and downtimes are tolerated
PROD	Environment with a definite quality of service (QoS) which will be offered to third-parties (e.g. SMEs and other research groups).

Sponsored production environment only during funding period. Afterwards all costs are routed through the customer's accounts (managed by the cloud provider. We do no billing!)

Sponsored usage of GPU/FPGA acceleration (for development and test envs) is limited to a certain amount of hours/month. For production the usage of those resources is routed through customer's own account.

AWS GPU-accelerated instances (e.g. p3.2xlarge) come with integrated Nvidia Tesla V100 GPU (~$11k)

AWS FPGA-accelerated instances (e.g. f1.2xlarge) come with integrated Xilinx UltraScale+ 64GB-DDR4 FPGA (~$7k)

AWS computing-optimized C5 instances offer the lowest price per vCPU and are ideal for running advanced compute-intensive workloads (e.g. theorem proving, model finding, ML training)

After the end of the funding period (01.2023) all development activities can stop, and consequently only testing and DevOps activities have been considered. All production costs can be routed through customer's own accounts

We consider no license costs since most/all relevant software infrastructure (Linux, Docker, OpenStack, Kubernetes, Hadoop/Spark, etc.) is open source

AWS costs calculator:

https://calculator.s3.amazonaws.com/index.html

AWS Educate Program:

https://aws.amazon.com/education/awseducate/

AWS Research Credits:

https://aws.amazon.com/research-credits/

Attachments

Letter of Support from the President of FU Berlin

Freie Universität Berlin

Der Präsident
Univ.-Prof. Dr. Günter M. Ziegler

Freie Universität Berlin, – Der Präsident -
Kaiserswerther Straße 16-18, 14195 Berlin

Kaiserswerther Straße 16-18
14195 Berlin

Telefon 49 30 838-73612
Fax 49 30 838-473612
E-Mail claudia.niggebruegge@fu-berlin.de
Internet www.fu-berlin.de
Bearb.-Zeichen VI C (k)
Bearbeiter/in Niggebrügge

21. Juni 2019

Unterstützung des Vorhabens *Explicit Normative Reasoning and Machine Ethics*

Sehr geehrte Damen und Herren,

mit Nachdruck unterstütze ich den von Herrn Prof. Benzmüller vorgelegten Antrag zur Einrichtung eines KI-Zukunftslabors zum Thema "Explicit Normative Reasoning and Machine Ethics (ENoRME)" an der Freien Universität Berlin.

Das dargestellte ENoRME-Projekt fokussiert auf unmittelbar bedeutsame und zukunftsrelevante Herausforderungen im Bereich der ethischen und rechtlichen Kontrolle von intelligenten autonomen Systemen. Es adressiert damit hochaktuelle Forderungen zur Entwicklung verantwortungsvoller Technologien im Bereich der Künstlichen Intelligenz. Dieses Thema zieht derzeit eine enorme Aufmerksamkeit auf sich.

Die Freie Universität Berlin bietet sich als ein hervorragend geeigneter und hochinteressierter Standort für das ENoRME Projekt an. Thematisch bieten sich erhebliche interdisziplinäre Anknüpfungspunkte zu anderen Aktivitäten an unserer Universität und der Berlin University Alliance, z.B. an den Exzellenzcluster MATH+ und den neuen interdisziplinären "Data Science" Studiengang mit Schwerpunkt am Fachbereich Mathematik und Informatik.

Herr Prof. Benzmüller ist ein außergewöhnlicher Wissenschaftler. Er hat ein hochkarätiges, interdisziplinäres Team von Wissenschaftler/innen und weiteren Projektpartner/innen zusammengestellt. Der Antrag stützt sich auf Vorarbeiten zur Automatisierung normativen Schließens, die Herr Prof. Benzmüller in enger Kooperation mit der Universität Luxembourg in den vergangenen drei Jahren erarbeitet hat. Diese bereits bestehende Kooperation zwischen der Freien Universität Berlin und der Universität Luxembourg soll im ENoRME-Projekt als Ausgangspunkt genutzt werden, um eine nachhaltige Infrastruktur und ein internationales Netzwerk zum Thema aufzubauen. Die Freie Universität Berlin begrüßt dieses Vorhaben nachdrücklich.

Mit freundlichen Grüßen

Prof. Dr. Günter M. Ziegler
Präsident

Letters of Intent

Participating Researchers: FU Berlin (Benzmüller, Rojas, Göhring), U Koblenz-Landau (Schon), U Luxembourg (Parent, Steen), U Bergen (Slavkovik), U Miami (Sutcliffe), U Liverpool (Dennis), U Campinas (Carnielli, Bueno-Soler), U Buenos Aires (Martinez)

Freie Universität Berlin

Fachbereich Mathematik und Informatik
- Das Dekanat -

Freie Universität Berlin, Fachbereich Mathematik und Informatik
Das Dekanat
Arnimallee 14, 14195 Berlin

Arnimallee 14
14195 Berlin

Telefon +49 30 838-54010
Fax +49 30 838-56746
E-Mail michael.weiss@fu-berlin.de
Internet www.mi.fu-berlin.de/
Bearb.-Zeichen We
Bearbeiter/in Dr. M. Weiß

Berlin, den 14.06.2019

**BMBF-Förderung von internationalen Zukunftslaboren in Deutschland zur Künstlichen Intelligenz:
Antrag von Herrn Prof. Benzmüller;
Projekt „Explicit Normative Reasoning and Machine Ethics (ENoRME)"**

Sehr geehrte Damen und Herren,

der Fachbereich Mathematik und Informatik unterstützt den Antrag von Herrn Prof. Christoph Benzmüller für das Projekt „Explicit Normative Reasoning and Machine Ethics (ENoRME)" nachdrücklich.

Für die Bearbeitung des oben adressierten Themas steht sowohl am Institut für Informatik und am Fachbereich Mathematik und Informatik wie aber auch an der gesamten Universität ein hervorragend ausgewiesenes und dynamisches wissenschaftliches Umfeld zur Verfügung.

Das Thema des Forschungsantrags bettet sich sehr gut in bestehende Strukturen der Freien Universität Berlin ein. Relevante Anknüpfungspunkte gibt es insbesondere zum Berliner Mathematik-Forschungszentrum MATH+ mit seinem besonderen Fokus auf datengetriebene Anwendungen von Künstlicher Intelligenz (KI). ENoRME adressiert flankierend Fragen der Maschinen-Ethik und der rechtlichen und ethischen Kontrolle von KI-Systeme; diese beiden Forschungsthemen ergänzen sich strategisch in idealer Weise. Als Mitglied von MATH+ und der assoziierten Graduiertenschule „Berlin Mathematical School (BMS)" ist Herr Kollege Benzmüller sehr gut positioniert, um als Bindeglied zwischen beiden Projekten zu fungieren.

Eine ähnliche gewinnbringende Situation wird sich für den neu eingerichteten Studiengang „Data Science" am Fachbereich ergeben (Start im Winter 2019 geplant). Auch eine geplante, interdisziplinäre Doktorandenschule an der Schnittstelle von „Geschichte und KI" mit dem John F. Kennedy Institut (ein Antrag wird derzeit von der Leiterin des Instituts, Prof. Gienow-Hecht, unter Mitarbeit von Prof. Benzmüller und Prof. Göhring vorbereitet) würde in signifikanter Weise von der Expertise und Betreuungskapazität der ENoRME-Gastwissenschaftler*innen profitieren.

Im Falle einer Bewilligung des Antrags von Prof. Benzmüller wird die Freie Universität Berlin als aufnehmende Institution Arbeitsplätze sowie die notwendige Grundausstattung zur Verfügung stellen. Ebenfalls wird die administrative und finanzielle Abwicklung des Projektes durch die Freie Universität Berlin gewährleistet.

Wir bestätigen ferner, dass es von Seiten des Fachbereiches keinerlei Bedenken zu der im Antrag dargestellten Mitarbeit von Prof. Benzmüller, Prof. Göhring und Prof. Rojas im Projekt ENoRME gibt und dass das Dekanat Freistellungsanträge der genannten Personen für die dargestellten Zeiträume (soweit erforderlich) befürworten wird.

Mit freundlichen Grüßen,

Univ.-Prof. Dr.-Ing. Rupert Klein
Dekan

Dr.-Ing. Andrea Bör
Kanzlerin