# Collaborative: EAGER: FDASS: A Law of Our Own: Towards An American Alternative to the European AI regulation

*Supreeth Shastri, Computer Science, University of North Texas*
*Mihailis Diamantis, College of Law, University of Iowa*

## 1 Introduction

**Overview and objectives.** Technology laws have emerged as effective tools to instill accountability in the use of technologies that have an oversized impact on the society. A prominent example is the European Union's General Data Protection Regulation (GDPR), which governs how organizations can store, share, and process personal data. Buoyed by the success of GDPR, the European Union recently proposed the Artificial Intelligence Act (AI Act) to mitigate the threats posed by a widespread use of AI. While there is a general consensus in society about the need for an AI regulation, the AI Act suffers from two critical issues: it is *legally rigid* and *computationally intractable*. Legal scholars, including PI Diamantis [10], have put forth a principled argument that its presumptive-fault framework and prescriptive design processes are unsuitable for regulating AI–they make technical innovations costly and iterative error reductions less likely. As the AI Act goes into effect, Europe is beginning to experience an exodus of AI companies, in turn stagnating AI innovation without necessarily making its society safe from AI harms. Worse yet, as seen in the case of GDPR, there is a tendency for other countries to clone European laws (the so called Brussels Effect [8]), which could severely affect the American AI landscape.

The overall objective of this proposal is to create a distinctly American framework for AI regulation that brings in accountability without impeding on innovation or our leadership in AI. Our central hypothesis is that rather than prohibit or censure any AI behavior outright, we can design an *objective-fault standard for AI* that compares AI's behavior with an external benchmark. Then, by tailoring the benchmark's standard to the social value and the current safety profile of the conduct at issue, the law can be made more or less demanding. The approach we propose is appealing, both legally and technologically, since (i) it always offers a reasonable baseline for comparing AI's fault, (ii) it incentivizes companies to release all and only AI products that increase safety, and (iii) as standards evolve, it demands more of AI. The genesis for this proposal comes from PI Diamantis' position paper on *Reasonable AI* [10] and subsequent discussions with PI Shastri on creating an AI regulation that is *AI-aware*. We plan to attain our objective by pursuing the following three aims:

1. *Laying the Legal Foundation for AI Negligence.* Our first goal is to build a functional legal framework for AI fault centered on a negligence standard. Since we do not intend to rely on "foreseeability" for determining AI's negligence, we will base it on tort law's standard of "reasonableness": an AI engaged in some task is reasonable if it causes less harm than the weighted average of harm that all actors, both AI and human, cause while engaged in that task. We will explore both the legislative and judicial options for translating our idea into an enforceable legal framework. Equally importantly, we will identify and analyze the key characteristics of this new AI negligence model both from a legal perspective (such as administrability and harm prevention) as well as a technology perspective (such as allowing continuous iterative progress and not hindering rapid innovation). The outcome of this phase will be an alternative AI regulation that is legally grounded and technologically aware.

2. *Developing AI Negligence Benchmarks.* In contrast to the first goal that focuses on the law, the second one targets its enforcement: designing a set of reference benchmarks to demonstrate the flexibility and generality of the law, and the ability of the stakeholders (namely, AI enforcement agencies, AI companies, AI consumers, and general public) to calibrate the implementation of the law. We aim to

develop negligence benchmarks for a representative set of AI application domains including transportation (self-driving vehicles), healthcare (disease diagnosis), and cybersecurity (detecting personally identifiable information). These tasks require engaging with experts from the respective domains in (i) characterizing the task, (ii) characterizing the harm, and (iii) quantifying the weighted average of harm per unit of task. We plan to open-source our benchmarks, methodology, and guidance on developing new benchmarks.

3. *Evaluating Against the AI Act.* Lastly, we aim to undertake a systematic comparison between the AI Act and our proposed standard by asking and answering three research questions: (i) does the law establish a measurable yardstick for enforcement, (ii) does the law increase overall AI safety now and over time, and (iii) does the law enable continuous and iterative advancement of AI. To ground our evaluation in real-world AI, we will limit our focus to general purpose AI systems that are commercially and publicly available. Then, by evaluating these systems across the representative AI applications identified in aim 2, our plan is to highlight the strengths and limitations of both approaches, and offer recommendations on choosing an AI regulation as a tool of accountability.

**Suitability for EAGER.** This work is particularly well suited for FDASS EAGER for several reasons. First, the project is interdisciplinary in scope (i.e, developing a law to bring accountability in AI systems) and radical in its approach (i.e., proposing to create a new AI-aware law from first principles); thus, fitting the category of *exploratory research* which employs integrative thinking of social and technical themes in designing accountable software systems. Second, by rejecting the status quo i.e., the AI Act and proposing to build an AI-aware alternative, our proposal has the *potential to transform* how accountable AI is developed and deployed in the U.S. and around the world. Finally, by achieving measurable accountability in AI without hampering AI innovation, our proposed work offers a *high payoff* to the society.

**Intellectual Merit.** This project will advance the state of knowledge in designing accountable software systems (DASS) with a focus on AI as the target software system, and technology regulation as the accountability mechanism. More specifically, our work will produce a *blueprint for an AI regulation* that avoids the tussle between law and AI, which is prevalent in the current state of the art, the European AI Act. We ground our findings by developing three AI negligence benchmarks, analyzing their legal- and computing characteristics, and proposing ways to implement this new AI standard in the U.S. legal system.

**PI preparation.** PIs Diamantis and Shastri are uniquely positioned to carry out the proposed work. Technology regulations are a key part of PIs' research interests and they are keenly aware of the emerging trends and challenges in the technology regulations landscape. For example, PI Diamantis has been developing philosophical and legal theories for regulating emerging technologies including AI [9, 10, 11]; PI Shastri has been studying the impact of GDPR on computing systems [17, 18] and tracking GDPR enforcement [19]. In recognition of their research contributions, PI Diamantis was invited to co-chair the Harvard Journal of Law & Technology symposium in 2023 as well as the Yale Journal of Law & Technology symposium in 2025, while PI Shastri was included in the European Data Protection Board's pool of experts in 2022.

# 2 Proposed Research

In this section, we describe our planned research activities, our approach to evaluating these tasks, and the expected outcomes.

## 2.1 Key Idea: A Negligence Standard for AI

Our approach to AI accountability is inspired by tort law's notion of reasonableness. The reasonable person standard is an objective standard that asks what a normal prudent person would have done in the same situation, as a benchmark to evaluate a defendant's conduct. While this is compelling, adapting the distinctly human notion of reasonableness to AI is challenging. This is because comparing AI to humans would set the bar too low for AI (since AI can and should outperform humans on many tasks). At the same time comparing AI to other AI presents its own challenges (since there are not enough AI in many contexts to establish a meaningful baseline). So, we define a hybrid negligence standard for AI as follows:

> **The AI Negligence Standard**: An AI engaged in some task is reasonable if it causes less harm than the weighted average of harm that all actors–both AI and human–cause while engaged in that task.

Based on this standard, when an unreasonable AI causes harm, it does so negligently. Rather than prohibiting any conduct outright (as in the case of presumptive-fault standards like the EU AI Act), objective fault standards compare the defendant's harmful behavior with some external benchmark. By making the benchmark more or less demanding, the law can can tailor the standard to the social value of the conduct at issue. Thus, applying the AI negligence standard to a particular AI undertaking a given task will follow three steps: (i) characterize the task and fix its unit of measurement, (ii) characterize the possible harms and fix their unit of measurement, and finally (iii) calculate the weighted average of harm per unit of task as exhibited by humans and AI.

**An illustrative example.** Consider the task of driving passenger cars in California. We choose the unit of measurement as the number of miles driven. Next, let us characterize harm as traffic accidents, which can be quantified by the number of accidents. From the California Department of Motor Vehicles 2022 report [16], we gather that self-driving cars have driven 5.7 million miles and humans have driven 310,823 million miles, with the former causing 150 accidents and the latter 57,734 accidents.

Average harm per unit of task is:
A = 150 / 5.7 = 26.31579 accidents per million miles for AI
H = 57734 / 310823 = 0.18575 accidents per million miles for humans

We compute the weight of each category as:
$W_A$ = 5.7 / (5.7 + 310,823) = 0.000018
$W_H$ = 310,823 / (5.7 + 310,823) = 0.999982

Then, the weighted average of AI and humans combined is:
Weighted Average = $(A * W_A) + (H * W_H)$ = 0.18622 accidents per million miles

This means that any AI that causes fewer than 0.18622 accidents per million miles is reasonable. For reference, Waymo, the leader in self-driving safety, had an average of 20 accidents per million miles [12]. By this yardstick, an accident caused by self-driving cars in 2022 would be deemed negligent. Next, to demonstrate the flexibility in adapting our benchmark to different situations, let us consider a different harm i.e., fatalities caused by traffic accidents. Using the same dataset as before, we find that self-driving cars had zero accidents involving death, whereas that number is 4,258 for human drivers. Applying the same weighting formula as before, we get the combined weighted average as 0.013786 fatalities per million miles driven—a bar that is exceeded by all present-day self-driving cars.

This formulation of AI negligence as an objective-fault standard is foundational for our proposal. However, to establish this as a viable alternative to the status quo, we must demonstrate its generality, its applicability beyond one example, and its advantages over the AI Act. In the rest of this section, we lay out three research aims that address these challenges.

## 2.2 Aim 1: Laying the Legal Foundation for AI Negligence

Our first research aim is to evolve the notion of AI negligence (as presented in Section 2.1) into a mature legal framework. Towards this goal we define two specific tasks, both of which will be led by PI Diamantis.

### 2.2.1 Research Plan

***Task 1: Translating AI Negligence into an Enforceable Mechanism.*** The goal of this task is to translate our legal theory into practice. We do so by exploring two orthogonal approaches: the first is a legislative option. We will create a blueprint for an AI regulation that is centered on the AI negligence standard and needs to be passed by a legislative body such as the U.S. Congress. This is the mechanism chosen for the AI Act, which was passed as a law by the EU parliament. The second option takes a judiciary path and demonstrates how the AI negligence standard could be integrated as an enforcement mechanism under the existing common law or torts. While the second option could be put into practice right away, its adoption will be piecemeal depending on the individual state judge decisions, whereas if one goes through a (likely lengthy and contentious) legislative process, it will likely have a broader scope and higher impact. We will explore the tradeoffs, limitations, and applicability of both approaches in the U.S. legal context.

***Task 2: Characterizing the Generality and Limitations of the AI Negligence Standard.*** If AI negligence is to be adapted as a viable standard, it is imperative to characterize its scope and limitations, both legally and technologically. For example, an AI that clears the reasonable threshold set by our standard may still cause harm and continue to do so until it falls below the threshold. This is due to an inherent tradeoff present in our standard that incentivizes safety only up to the point where a reasonable baseline is met. As we design and develop AI negligence benchmarks for a broad range of domains (in Section 2.3), we will reevaluate the generality and limitations of our framework. Similarly, comparing AI negligence to the AI Act (in Section 2.4), we will identify its strengths and limitations more accurately.

Another important dimension to explore is the impact of the AI negligence standard on society. For instance, as AI improves its baseline performance, it could also raise the legal standard for human performance. While this may already be happening at an abstract level, by succinctly quantifying the AI superiority, our standard may put human jobs and roles in jeopardy. We believe it is important to capture both direct and second-order impacts of regulations.

### 2.2.2 Evaluation Plan

We will solicit feedback from legal scholars (via peer-review process and conference presentations), policy makers (via ad hoc engagements), and the general public (via open-sourcing) on the quality and correctness of our work. We will undertake both qualitative and quantitative evaluations based on these inputs.

### 2.2.3 Expected Outcomes

We expect this aim to result in two novel contributions: (i) a blueprint for an alternative AI regulation that is legally grounded and technologically aware, and (ii) two pathways for implementing AI negligence in practice. We also expect to engage with interested federal and state agencies as well as advocacy groups to share our approach for AI regulation.

## 2.3 Aim 2: Developing AI Negligence Benchmarks

Our second aim focuses on the practical enforcement of the proposed law by designing a set of reference benchmarks. We choose a representative set of AI domains to apply the negligence standard, and to demonstrate its flexibility and generality. This task will be led by PI Shastri.

### 2.3.1 Research Plan

***Task 3: Developing a Comprehensive Benchmark for Self-Driving Vehicles.*** The goal of this task is to create a comprehensive benchmark for self-driving vehicles that supports a broad variety of task- and harm definitions, that is informed by empirical data from transportation bodies, and that allows for extensibility as new scenarios emerge. We do so along three axes:

- *Benchmark metrics*: we will incorporate well-established driving harms of death, bodily injury, property damage as well as emerging psychological harms from phenomena such as phantom braking, cyber attacks, loss of control, among others. We also intend to develop metrics to track failures that are systemic in nature (which are akin to say, drunken driving for humans) which should lower the threshold for negligence and/or incur higher penalties for the same harm.
- *Benchmark tracks*: We will define the operating contexts of vehicles as "tracks". This allows us to distinguish the performance of driving vehicles under (i) different weather conditions like snow, rain, wind, heat etc, (ii) different states and terrains, (iii) different generations of technology for sensing, learning, etc.
- *Benchmark data*: The real strength of an objective standard comes from underlying empirical data, and to that effect, we will build an open-source and publicly accessible traffic data repository. We will develop software systems that automatically crawl data from federal institutes on traffic and safety as well as regional departments of motor vehicles, and publish the state of the art benchmark results on our project website.

We will ground our work by engaging with experts at UNT Autonomous Transportation Institute, the largest institute of autonomous vehicle research in Texas as well as seek feedback from safety researchers of the leading self-driving AI companies.

***Task 4: Establishing AI Negligence Benchmarks for Healthcare and Cybersecurity.*** Our next task focuses on generalizing our framework with the development of two novel benchmarks. The choice of healthcare and cybersecurity is driven by their divergence in types of AI harms, their categorization into different risk brackets under the AI Act, and the availability of expertise at UNT and UI. In healthcare, we will focus on medical chart summarization [2] and patient session scribing [5]—activities for which the UI Healthcare has rolled out advanced AI systems in 2024 [6]. In cybersecurity, we will focus on the detection and removal of Personally Identifiable Information (PII) from data storage systems—a challenge that has gained prominence in the GDPR era [1, 7].

### 2.3.2 Evaluation Plan

We will evaluate the success of our tasks using the following metrics: (i) the ability of the AI negligence framework to accommodate domain-specific characterizations of harm, (ii) feedback from domain experts about the quality and meaningfulness of the benchmark, and (iii) how these benchmarks strengthen or evolve the properties of the AI negligence framework. We will evaluate our findings both qualitatively and quantitatively based on these metrics.

### 2.3.3 Expected Outcomes

We expect this aim to result in the creation of three novel negligence benchmarks covering a broad spectrum of AI application domains. We will open-source our methodology, datasets, and software systems for the three domains, while also providing guidance on extending this work to newer domains. We anticipate publishing our work in an interdisciplinary journal.

## 2.4 Aim 3: Evaluating Against the AI Act

Our final aim is to undertake a comprehensive comparison of the AI Act and our AI negligence standard by asking and answering three research questions: (i) does it establish a measurable yardstick for enforcement, (ii) does it increase the overall AI safety now and over time, and (iii) does it enable continuous and iterative advancement of AI. We do so with two specific research tasks, which will be co-led by both the PIs.

### 2.4.1 Research Plan

***Task 5: Applying the AI Act to Case Studies.*** To make our comparison fair and in-depth, we apply the AI Act to the same three representative AI applications identified in aim 2. Our initial investigation informs us that autonomous vehicles and healthcare come under the *high-risk* category of the AI Act, whereas cybersecurity falls under the *limited-risk* category. We believe this is a good mix since the AI Act is largely focused on these two categories while outrightly prohibiting the *unacceptable-risk* category and leaving the *minimal-risk* category unregulated. We will perform both qualitative analysis and empirical evaluation with a goal to highlight the strengths and limitations of the AI Act. Our prior work on analyzing the impact of GDPR on database systems [17] serves as a useful template for this task.

***Task 6: Systemic Comparison.*** Finally, we zoom out to perform a systemic comparison of the entire regulation. We identify four key criteria for this task by considering both the legal requirements and AI specific characteristics. Table 1 provides a concise summary of these criteria. This formulation builds on PI Diamantis's prior work on Reasonable AI [10], and provides us with a clear set of rubrics to evaluate any AI regulation including ours.

### 2.4.2 Evaluation Plan

To evaluate whether our tasks are effective against the stated research goal, we will assess our outcomes by two metrics: (i) *Coverage of the AI Act* i.e., how many articles of AI Act are exercised during our analysis, and (ii) *Completeness of the rubrics* i.e., do the rubrics cover all aspects of the existing and proposed AI regulations. Also, we will seek feedback from policy experts and through peer review process to ground our research.

### 2.4.3 Expected Outcomes

We expect these tasks to result in two novel contributions: (i) compliance profiles for three real-world AI applications for the AI Act, and (ii) a methodology for systemic assessment of AI regulations. We anticipate our work to result in a full-length paper in a top computing- or inter-disciplinary journal.

| Criteria | What it measures |
|---|---|
| Administrability | Does the standard offer a clear and predictable threshold for AI's behavior? |
| Progress | Does the standard become more demanding as AI's capabilities improve? |
| Safety | Does it incentivize creating AI that improves safety, and disincentivize AI that decreases safety? |
| Innovation | Are the liability thresholds flexible enough to not adversely impact the cost of innovation? |

**Table 1:** *Four key criteria for assessing the effectiveness of AI regulations.*

# 3 Broader Impacts

This project address an emerging socio-technical problem and will lead to positive impact on research, education, and society. In particular, our contributions align with three of the broader impact criteria identified in the NSF solicitation:

**Increased economic competitiveness of the U.S.** There is an ongoing tussle in our society about AI's oversized influence on people's life and how people can have a say in its evolution. Though there is some acknowledgement that AI needs to be accountable, there is also a concern that a heavy handed European-style regulation may adversely affect the American leadership in AI. This project—by demonstrating a legal framework that brings accountability to AI without impeding its innovation—completely eliminates this societal tussle. We believe that a successful execution of this project will pave the way for a meaningful and well-accepted AI regulation that is centered on the American ethos.

**Increased partnerships between academia, industry, and others.** The scope and timeliness of this project provides a unique opportunity to reach out to a broad community of policy makers, regulators, and industry leaders. More specifically, the PIs plan to pursue the following:

- *Partnering with industry and policy makers*. PI Diamantis has a successful track record on this. In March 2023, he co-organized *Beyond the FTC: The Future of Privacy Enforcement* [14], a multi-disciplinary symposium at the Harvard University that hosted the FTC Director Samuel Levin, U.S. Congresswoman Lori Trahan, and EFF Director Cindy Cohn, among others. In March 2025, he co-organized a symposium on *Governing Data* [15] at Yale University that brought together the Chief AI Officer of the US DOJ, global policy leaders of Meta, and 170 other scholars from various disciplines. PI Diamantis plans to organize a similar event with focus on AI governance.
- *Real-world adoption of our research artifacts*. The PIs are committed to making all our datasets and software systems easily accessible and releasing them as open-source and free to all. This commitment is evident in our prior work such as GDPRbench [3] and GDPRxiv [4], both of which have fully-featured websites and fully-functional open-source tools.

**Development of a globally competitive STEM workforce.** Our research objectives enable the students to develop skills that are global in nature and interdisciplinary in practice. The project will enable the PIs to achieve the following educational goals:

- *Enhancing AI education at UNT*. PI Shastri will integrate the findings of this research into his graduate course *AI and Data Governance*, which will be offered annually starting Spring 2026. The UNT computer science department intends to make this a mandatory course for graduate students in the MS AI program (a cohort of 300+ students).
- *Training interdisciplinary scholars*. The PIs are committed to individually training and mentoring $\sim$20 students across the UI law college and UNT CS department over the next two years (both undergraduates and graduates, many not directly funded by this project). We expect these students to become leading researchers and practitioners of the lawful computing paradigm.

# 4 Anticipated Timeline

This proposal is a significant undertaking by the PIs, comprising of both interdisciplinary research and collaborative activities. We will bootstrap the project in October 2025 and organize our work over the next

| Aims | 2025 Fa | 2026 Sp | 2026 Su | 2026 Fa | 2027 Sp | 2027 Su |
|---|---|---|---|---|---|---|
| **Aim-1: Legal Foundation** | | | | | | |
| Enforcement mechanisms | | | | | | |
| Generalizing AI negligence | | | | | | |
| **Aim-2: Negligence Benchmarks** | | | | | | |
| Self-driving vehicles | | | | | | |
| Healthcare | | | | | | |
| Cybersecurity | | | | | | |
| **Aim-3: Compare w/ AI Act** | | | | | | |
| Apply AI Act to case studies | | | | | | |
| Systemic Comparison | | | | | | |

**Table 2:** *Estimated timeline for project deliverables and events.*

two years as shown in Table 2. The tasks that focus on individual case studies are colored for ease of tracking, while the tasks that are broad and systemic in nature are painted in black. We plan to complete the tasks related to self-driving vehicles (both the development of AI negligence benchmark and the evaluation against AI Act) by the end of Spring 2026. Similarly, we will complete the tasks related to Healthcare by Fall 2026, and those corresponding to cybersecurity by Spring 2027. On the legal foundation aim, we will work on the mechanisms to enforce AI negligence in Spring and Summer of 2026, and take up our effort to generalize our new standard in Spring 2027. We will wrap up our project by performing a systemic evaluation of AI negligence standard against the AI Act in Summer 2027.

# 5 Results from Prior NSF Support

*(Award #2335659) Education DCL: EAGER: Taking Law into Our Own Hands: Infusing Law into CS Education, Early and Often.* **PIs.** Rishab Nithyanand and *Mihailis Diamantis*. **Years.** 2023-2025. **Amount.** $299,994.

*Intellectual merit.* Co-leading an interdisciplinary group of law and CS students, PIs Nithyanand and Diamantis produced a new dataset which contains expert annotations of privacy policies, specifically focusing on organizations' response to tech regulation mandates. This dataset and the associated publication appeared at EMNLP 2024 [13]. Another dataset project related to creating a taxonomy of privacy policy terms is underway.

*Broader impacts.* Co-PI Diamantis co-organized a symposium about data governance at Yale Law School in March 2025 with over 170 people participating. It will result in publication of a 14-essay issue of the Yale Journal of Law and Technology. Additionally, co-PI Diamantis and PI Nithyanand have redesigned their Law/CS course "Privacy Law & Technology" to create competitions related to auditing regulatory compliance. CS and Law students collaborate to design and explain the significance of privacy audit interventions.

PI Shastri has not received any funding from the NSF yet. It must be noted that, in 2023, PI Shastri was part of the team that was awarded the aforementioned NSF grant #2335659. However, before the project could start, PI Shastri left the country for an overseas academic position and had to withdraw from the NSF grant. Thereafter, he could not contribute to that project, nor did he or his students receive any funding from the NSF.