Q1 2022

Ethical AI Ecosystem

00. TABLE OF CONTENTS

EAIDB

01. INTRODUCTION

The demand for ethical AI services (including terms like "explainable AI" or "responsible AI") has skyrocketed, in part due to some of the troubling practices employed by large technology companies. Everyday media is full of news of privacy breaches, algorithmic biases, and AI oversights. In the past decade or so, public perceptions have shifted from a state of general obliviousness to a growing recognition that AI technologies and the massive amounts of data that power them pose very real threats to privacy, to accountability and transparency, and to an equitable society.

The Ethical AI Database project (EAIDB) [1] seeks to generate another fundamental shift - this time from awareness of the challenges to education of potential solutions. We attempt to spotlight a nascent and otherwise opaque ecosystem of startups that are geared towards shifting the arc of AI innovation towards ethical best practices, transparency and accountability. EAIDB, developed in partnership with the Ethical AI Governance Group, presents an in-depth market study of a burgeoning ethical AI start-up landscape geared towards the adoption of responsible AI development, deployment and governance. We identify five subcategories, then discuss key trends and dynamics.

Primary Report Contact: Abhinav Raghunathan (abhirags@seas.upenn.edu)

EAIDB

02. MOTIVATION

Ethical AI is quickly gaining prominence for various stakeholders across the innovation landscape-from startup executives developing AI-first solutions and the investors who fund them to enterprise customers deploying them and society at large. Policy is beginning to make an appearance on a global scale in regards to safe, responsible AI. Ethical AI is emerging as a ubiquitous need for companies-no matter what the problem statement might be. The sheer volume of companies identified as "ethical AI companies" corroborates this reality.

We define an "ethical AI company" as one that either provides tools to make existing AI systems ethical or one that builds products that remediate elements of bias, unfairness, or "unethical-ness" in society. The number of such companies has shown significant growth throughout the last five years.

The motivation behind this market research is multidimensional:

1 Investors seek to assess AI risk as part of their comprehensive profiling of AI companies. EAIDB provides transparency on the players working to make AI safer and more responsible.

2 Internal risk and compliance teams need to operationalize, quantify and manage AI risk. Identifying a toolset to do so is critical. There is also an increasing demand for ethical AI practices.

3 As regulators concretize policy around ethical AI practices, the companies on this list will only grow in salience. They fundamentally provide solutions to the problems AI has created.

4 On a more philosophical note, AI should work for everyone, not just one portion of the population. Enforcing fairness and transparency in black-box algorithms and opaque AI systems is of the utmost importance.
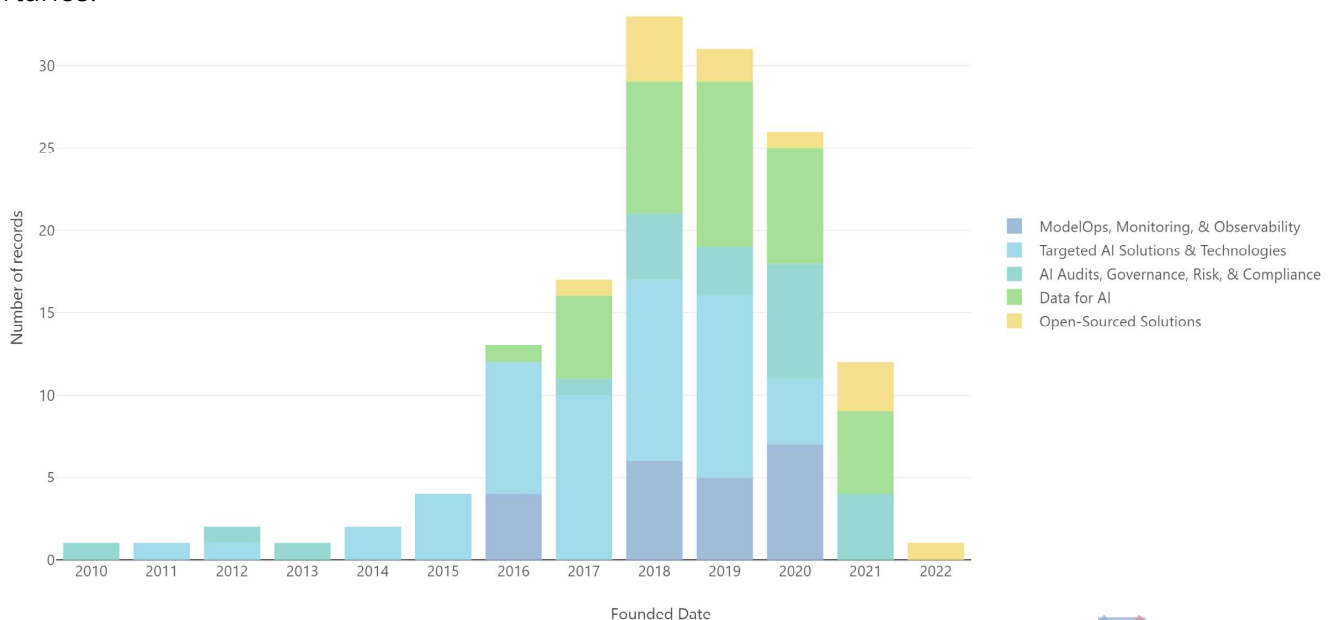
03. METHODOLOGY

1 **We place priority on recent, early-stage startups founded post-2015.**

Older companies have a tendency to add ethical AI principles in a post-hoc way. The companies in our database and in our market map have truly innovated. Some exceptions can be made if a company is truly ahead of its time and does great work within the Ethical AI ecosystem.

2 **Cybersecurity, climate tech, and related fields are important, but outside of our scope.**

Our focus is on companies that affect issues in society-level ethics like lack of representation, discrimination, accountability, etc. Climate, for example, is much more than just a society-level issue.

3 **We classify companies based on their main line of business.**

Our categories are designed to represent several vertical and horizontal business models (see the ecosystem below). Some companies might have products that belong in multiple categories.

4 **Companies must have a specialized concentration on fairness / bias mitigation.**

For example, there are many model management and quality assurance platforms, but those that do not have specific features that allow for bias detection / monitoring fall outside of our scope. AI quality assurance is by itself indirectly ethical, of course, but there needs to be a recognition of fairness / bias concerns and some focus on how / why a platform can help.

EAIDB

04. DATA FOR AI

DESCRIPTION

Companies in this category provide specific services to maintain data privacy, detect data bias early, or provide alternative methods for data collection / generation to avoid bias amplification later in the machine learning lifecycle. A large portion of companies in this category specialize in synthetic data: essentially, generating a new, artificial dataset that is statistically guaranteed to behave similarly to the old one. However, because this data no longer refers to real people or true information, there are no privacy concerns. These companies compete on how similar their synthetic sets are relative to the original datasets and how flexible their solutions are (for example, can the product handle both unlabeled and labeled data?). Other subcategories include data sharing (data permissions, safe transfer, etc.), data privacy (anonymization, differential privacy, etc.), and data sourcing (representative samples, minority amplification, etc.). Companies generally deliver their services through APIs or CLIs.

TRENDS & DYNAMICS

Some of the more interesting players in this space are experts at one particular kind of synthetic data. Datagen [2], for example, focuses primarily on facial data by allowing datasets to contain diverse skin colors, hairstyles, genders, angles, etc. to minimize the risk of biased facial recognition technologies. Data observability and sourcing platforms like Galileo [3] or Snorkel AI [4] offer products that ensure data quality - the former automatically correcting bias in unstructured data and the latter performing fair, automatic labeling with data versioning and auditing services.

Given the importance of data in the quality of AI systems (garbage in, garbage out), "Data for AI" companies will only grow in importance. Versioning and auditing in the context of bias will continue to play a large role in basic offerings provided by ethical AI companies. It might be that synthetic data companies will crowd each other out, since barriers to entry are fairly low and only a few have superior products or have adopted niche data areas (vision, text, etc.). Data sharing and privacy companies overlap to some degree with cybersecurity and secure computing, which are their main sources of competition. What lacks in this space is "context-conscious data mining," or a mining / sourcing platform that understands the context of the dataset and then assess potential bias concerns.

EAIDB

05. MODELOPS, MONITORING, & OBSERVABILITY

DESCRIPTION

Members of this category provide specific tooling to monitor and detect prediction bias (however, this may be defined in context). Usually self-defined as "quality assurance for ML," these companies specialize in black box explainability, continuous distribution monitoring, and multi-metric bias detection. For the purposes of this study, MLOps companies that provide generic monitoring services fall outside of scope.

Companies in this space are somewhat uniform in their offerings. MLOps platforms like Fiddler [5] or Arthur [6] have features such as drift detection and bias detection / monitoring, but also touch on explainability (which is relatively low-effort to add in). Others in this space focus specifically on deep learning visibility. Many companies package themselves as traditional MLOps companies with the added upside of bias-related software. One increasingly interesting subtopic in the MLOps world is model versioning - which is very closely related to data versioning and overlaps with the next category, governance.

TRENDS & DYNAMICS

The product differentiation between top players in this category is relatively small. The incumbents in the MLOps space (Amazon, DataRobot, etc.) have quickly adopted bias-related technologies that present some form of competition, but specialized firms like those in this category surely offer better integration and better monitoring products because these are their primary mode of business. Many constituents in this space have their own internal labs in which theory is applied to practice. This is rare in the corporate world, but the companies that constantly pay attention to the latest methods of bias detection and resolution and put them into use will surely cement themselves at the top.

Explainability in deep learning is a debatable proposition in its own right. Shapley values or LIME are disputed as proper explanatory methods. In regards to data involving humans, the use of deep learning may decline because these models present low interpretability and are therefore riskier investments. Regulatory compliance is, therefore, significantly more difficult. In this case, deep learning models become less desirable and platforms that focus on deep learning explainability like XaiPient [7] become, by extension, less desirable. The theory around visibility in deep learning models will inform whether this subspace will wax or wane in demand.

EAIDB

06. AI AUDITS, GOVERNANCE, RISK, & COMPLIANCE

DESCRIPTION

Members of this group are usually specialist consulting firms or platforms that establish accountability / governance, quantify model and / or business risk, or simplify compliance for internal teams within AI systems. Consulting firms differentiate themselves according to experience and specialization (ex. NLP, deep learning, etc.). Some, like BNH.ai [8], are law firms that specialize in bias consulting. Top consulting firms (BCG, McKinsey, etc.) are main competitors and generally attract the most diverse clientele, though as larger and older companies they have different methodologies and expertise than smaller, more recent players.

Software platforms tend to focus instead on increasing accountability and transparency by making AI models shareable with all stakeholders. Some allow for the "automation" of governance such that risk is visible in the same way throughout the ML lifecycle. Documentation, reporting, and other features are usually included alongside the main product. Certain niche companies, like EthicsGrade [9], apply objective frameworks to companies and assign "grades" to boost transparency among consumers, investors, and regulators alike. Automated compliance is also a subcategory, in which companies like Relyance AI [10] enforce contractual compliance at the code level.

TRENDS & DYNAMICS

The growth in AI Audits & GRC companies has been constant over the last five years, thriving in part due to weaker policy and indecision among international governments regarding what best practices in AI should be. These firms are naturally the most flexible companies in the EAIDB and are able to assist with various parts of the AI / ML lifecycle in a context-conscious way - something that platforms will always struggle with. However, as better metrics are created and the demand for automatic ML compliance increases, platforms like Credo AI [11] may begin eroding firms' market shares. Automated GRC is still very general - there is still lots of room for specialized, bias conscious GRC solutions.

EAIDB
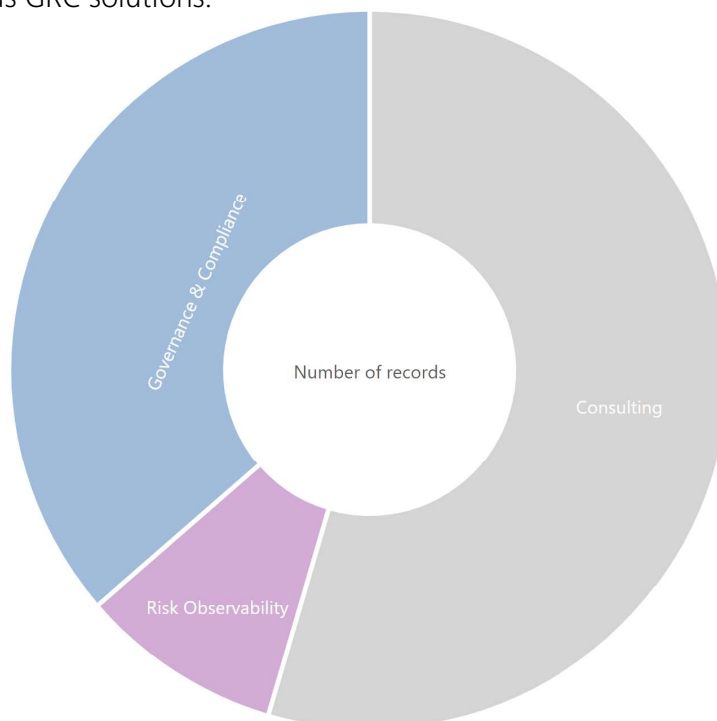
07. TARGETED AI SOLUTIONS & TECHNOLOGIES

DESCRIPTION

Targeted AI solutions encompasses AI companies that attempt to solve a particular ethical issue in a vertical. Sometimes describable as "a more ethical way to __," these companies are usually contained within labels like hiretech, insuretech, fintech, healthtech, etc. Targeted AI technologies refers more to general AI that is more horizontally-integrated and vertically-applicable. Some examples of common horizontals are toxic content detection, deepfake detection, and ethical facial recognition.

Due to the "hiring bias boom" of the mid-2010s, a majority of the companies in this category are hiretech companies. Pave [12] is a benchmarking startup meant to provide tools to solve wage inequality while Diversio [13] can quantify and help improve DEI initiatives. Fair treatment in lending with companies like FairPlay [14] is also a hot topic (though fairness in finance has always been more regulated than in other fields). Other extremely niche companies like Flock Safety [15] and Citibeats [16] touch on very interesting use cases for responsible AI design and enjoy relatively few competitors.

TRENDS & DYNAMICS

As applications where ethical AI is necessary continue to increase, targeted AI solutions will only become more popular. With the advent of "deepfakes," for example, specialized companies involved with the ethical implications of the technology (like Sentinel [17] for identity theft) have been developed. The recent completion of the deciphering of the human genome might spur more AI companies related to biodata and genetic construction. A longer-term application space within healthtech might be related to ethical concerns regarding this genetic data (not limited to privacy). There may be a new wave of insurance companies under insuretech that use alternative methods of calculating prices. Just Insure [18], for example, uses a customer's actual driving to calculate what they should pay. This removes the need for background checks involving other types of data (like race, for example) and therefore may reduce proxy bias. There is a lot of room for very niche companies to grow in this category - there are verticals (such as ethical crime detection or ethical social media analytics) in which there is almost no competition.
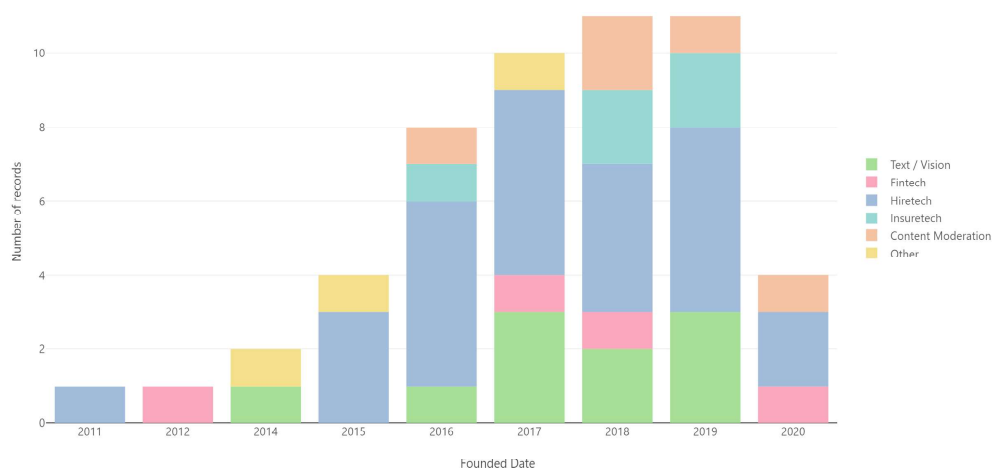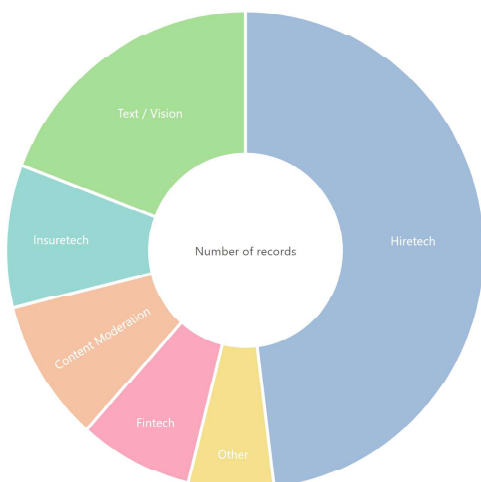
08. OPEN-SOURCE TECHNOLOGIES

DESCRIPTION

As the name suggests, this category contains fully open-source solutions meant to provide easy access to ethical technologies and responsible AI. The companies behind these frameworks are not usually for-profit (though some, like Credo AI's Credo Lens [19] or IBM's AI Fairness 360 [20], are), but their open-source technology is usually a good approximation of where the cutting-edge of applied ethical AI research is. open-source tools certainly play their own role within the startup ecosystem because they provide access to cheap tools that other, non-specialist firms can employ. Most open-source tools are concerned with privacy, bias detection, and explainability. The shortcomings with this category are consistent with shortcomings of open-source in general: vulnerability to malicious users, lack of user-friendliness, and lack of extensive support systems. However, they provide a baseline that companies need to constantly beat by providing more flexibility, better support, and easier access to warrant the prices they charge.

TRENDS & DYNAMICS

Companies usually cannot keep up with the rapid pace of theoretical development in a field as dynamic as algorithmic fairness. Open-source frameworks can because the barrier to creating a GitHub repository and starting an open-source project is next to nothing. This category will always continue to grow because there are no "competitors." Deepchecks [21], for example, is an open-source framework for ML model testing, with the ability to write tests for bias in vision and standard models. Its code repository boasts a high activity rate of three commits per day. However, the open-source community is generally a good place to establish a baseline and identify aspects of ethical AI that are lacking from the for-profit world. Through just open-source frameworks, one can generate a synthetic dataset (which is privacy-preserving and compliant), enforce fairness in ML training, explain these models, then audit them for proxy bias. Products in the for-profit world must beat open-source frameworks in scale, speed, etc.

EAIDB

09. ECOSYSTEM TRENDS

1 **Automated GRC and MLOps will dominate consulting firms.**
As policy is created, refined, and further defined, consulting firms may decline in popularity and MLOps / GRC platforms may rise due to the ability to programmatically enforce compliance given concrete metrics. This is counter to the trend over the last ten years in which consulting firms outpaced automated solutions.

**2** **Incumbents will try to replicate newer Ethical tech in-house.**
Incumbents in the space will start incorporating less effective versions of bias-related technology in an effort to keep their platforms viable in this new, bias-conscious world. They lack the specialty, expertise, and first-mover advantage of these startups but do have a well-established client base to draw from. Whether these companies can tap into new markets as effectively remains to be seen.

3 **The "modern tech stack" will shift to an "ethical tech stack."**
The modern "tech stack" will quickly evolve into an "ethics stack" in which the flow of data is carefully monitored, documented, and analyzed through products provided by companies in the aforementioned categories. For example, a company might employ Gretel [23] for data privacy and synthetic data, Akira [24] for MLOps management and continuous bias detection, then Saidot [25] for model versioning and governance.

4 **Demand for Ethical AI solutions will perpetually increase.**
Performing AI services correctly is domain-specific, context-specific, and very unstable (i.e. always needs to be monitored, checked for quality, etc.). There's no "one-and-done" approach to AI. The "boom" for ethical AI is estimated to be somewhere from the mid- to late-2020s and will follow a curve similar to "ethics predecessors" like cybersecurity in the late-2000s and privacy in the late-2010s. There will be a time in which policy, real case studies of AI gone wrong, and new discovery of biased AI (and a genuine desire to fix it) nudges companies in the right direction (through fear or will), leading to large demand and more inclusions in EAIDB.
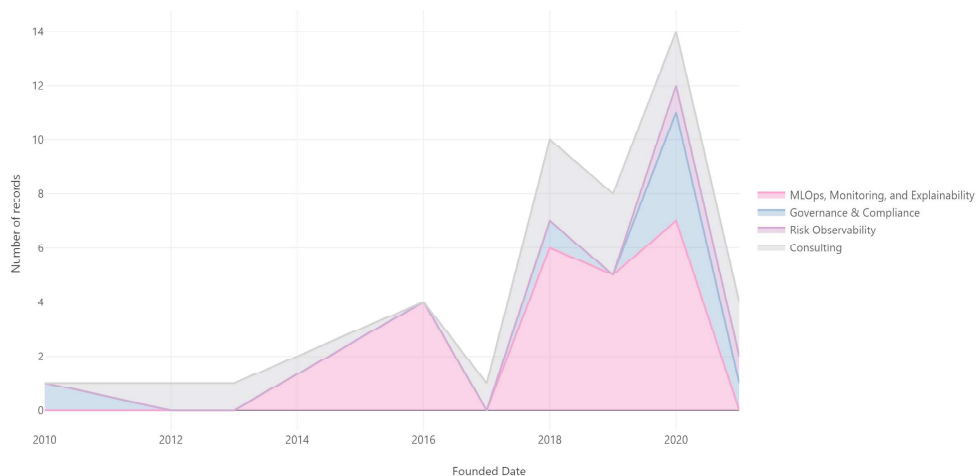
10. CONCLUSION

This is a nascent ecosystem, but it is growing rapidly and is expected to increase in momentum as motivations improve. Incipient measures to track and measure this area will certainly increase in turn. EAIDB reports will be published on a quarterly basis to lift the veil and spotlight both the importance and growth of this space. Over time, trend lines will emerge and taxonomies will shift to adapt to the dynamic reality of this ecosystem. In the meantime, we hope this report sheds some light on what is undoubtedly a fascinating and critical area of innovation.

Abhinav Raghunathan

EAIDB

11. APPENDIX

Companies in this Report

[1] Ethical AI Database // Public, vetted database of AI startups providing ethical services (eaidb.org).

[2] Datagen // Synthetic data for computer vision teams (datagen.tech).

[3] Galileo // Next generation data toolchain for ML developers (rungalileo.io).

[4] Snorkel AI // Data-centric AI for the enterprise (snorkel.ai).

[5] Fiddler AI // Machine learning monitoring for enterprise (fiddler.ai).

[6] Arthur AI // #1 Monitoring platform to manage ML (arthur.ai).

[7] XaiPient // Intelligence layer for Martech data (xaipient.com).

[8] BNH.ai // A boutique law firm focused on AI and analytics (bnh.ai).

[9] EthicsGrade // Stakeholder-centric ESG data company (ethicsgrade.io)

[10] Relyance AI // Manage privacy, data governance, and compliance operations seamlessly, all on a single, intuitive platform (relyance.ai).

[11] Credo AI // Empowering organizations to create AI with the highest ethical standard (credo.ai).

[12] Pave // Plan, communicate, and benchmark your compensation in real-time (pave.com).

[13] Diversio // People intelligence platform that measures, tracks, and improves diversity and inclusion (diversio.com).

[14] FairPlay AI // Assess your automated decisioning models in minutes to increase both fairness and profits (fair-play.ai).

[15] Flock Safety // Public safety operating system that helps communities and law enforcement in 1200 cities to eliminate crime (flocksafety.com).

[16] Citibeats // Today's societyal changes through real-time data analysis (citibeats.com).

[17] Sentinel // Defending against deepfakes and information warfare (thesentinel.ai).

[18] Just Insure // Data driven, pre-paid, and pay-per-mile car insurance (just.insure).

[19] Credo Lens // AI assessment framework (credoai-lens.readthedocs.io).

[20] IBM AI Fairness 360 // Examine, report, and mitigate discrimination and bias in machine learning models throughout the AI application lifecycle (aif360.mybluemix.net).

[21] Deepchecks // Build test suites for ML models and data (deepchecks.com).