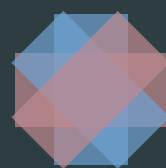


Q2 2022

Ethical AI Ecosystem



EAIDB

TABLE OF CONTENTS

- Introduction.....01
- Motivation.....02
- Methodology.....03
- Additions & Deletions.....04
- Investing in Ethical AI
 - Venture Funding Summary.....05
 - Top Investment Firms in Ethical AI.....06
- Ecosystem Trends
 - Data for AI.....07
 - MLOps, Monitoring, and Explainability.....09
 - AI Audits, Governance, Risk, and Compliance.....10
 - Targeted Solutions and Technologies.....11
- Barriers to Ethical AI
 - Lack of Cultural Diversity.....12
 - Demand Distribution for Ethical AI.....14
- EAIDB Initiatives & Partnerships.....15
- Conclusion.....16
- Appendix.....17

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 startup 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)

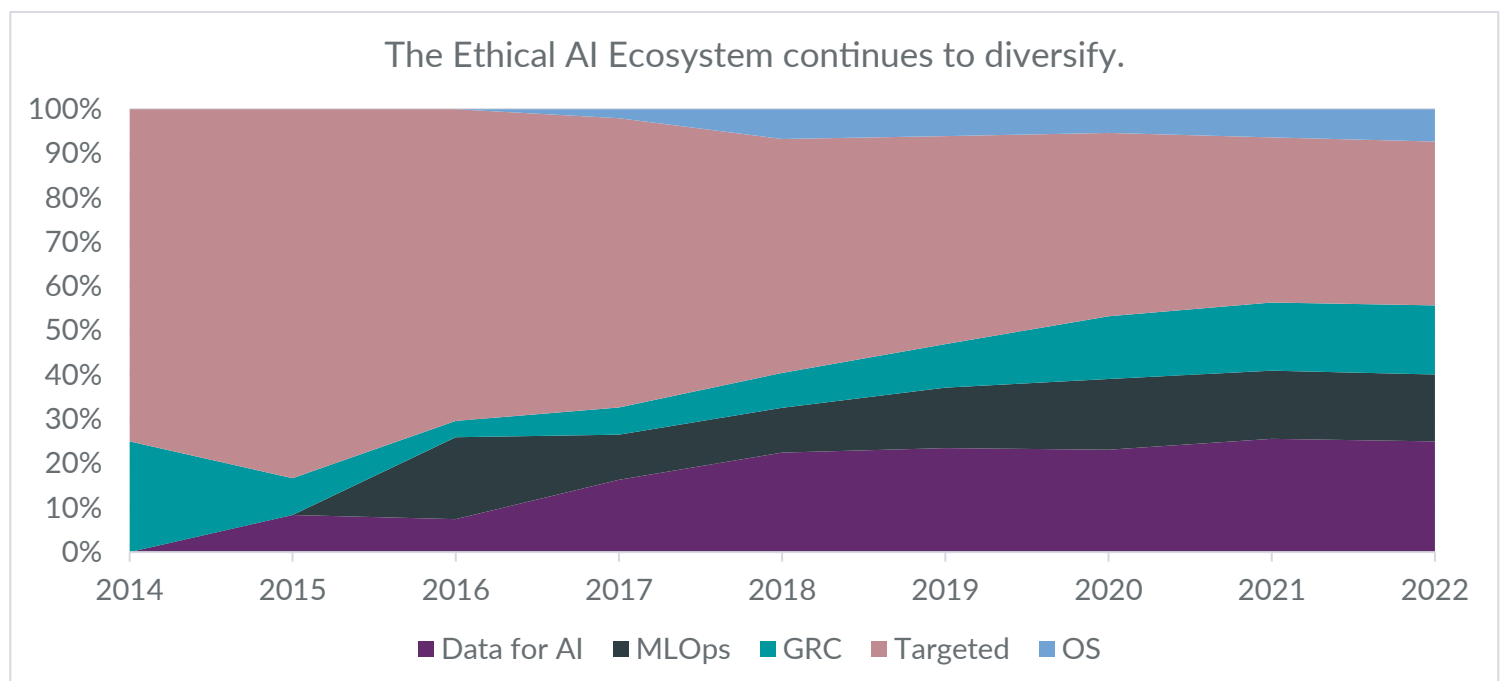
Motivation

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 "unethicalness" 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.



Methodology

1 **We place priority on recent, early-stage startups founded post-2014.**

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 are pioneers who have truly innovated within the space. 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.

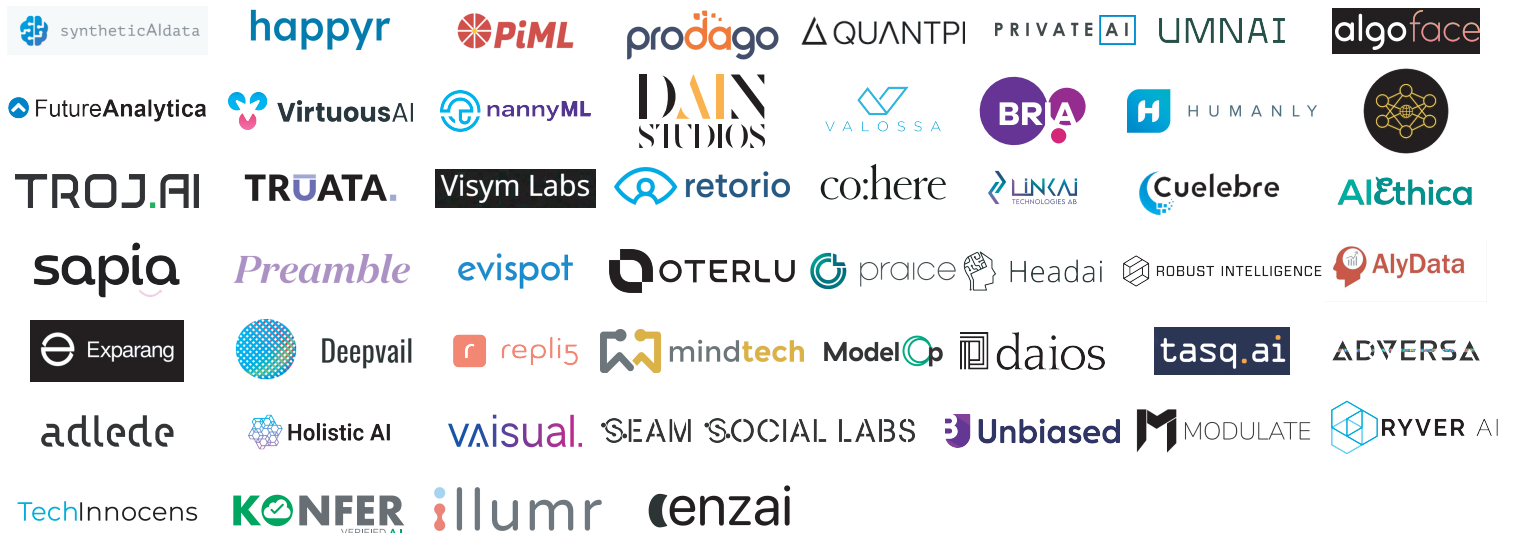
4 **Companies must have a specialized concentration on fairness, bias mitigation, or privacy.**

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.

New Additions & Deletions

ADDITIONS

Since the release of the Q1 2022 report, EAIDB has continued market research and has identified **52 new companies** that provide value to the ethical AI ecosystem. This is an increase of about 32% since our last report. To all the below organizations, thank you for making a difference and welcome to EAIDB!



DELETIONS

We've also identified some companies in EAIDB that we feel no longer fit our criteria upon closer inspection. Some are simply too matured (we have a strong preference for startups that are fairly early on in their journey, Series E or earlier), some are inactive, and some we feel do not directly relate to AI ethics or its many subcategories. As we continue to refine our criteria, EAIDB will always be transparent in the changes made to its map and constituents. The following companies were removed from the database:

Upstart went public in 2020 and is not an early-stage company.

Tweedle is inactive.

nuAI is inactive.

Chatterbox Labs was founded in 2010 - well before the scope of this map.

Darwin AI has a product that does not directly interact with humans or society.



Investing in Ethics

2022 Funding Highlights
Top Investment Firms in Ethical AI

The first half of 2022 saw over \$960m in funding from 39 Ethical AI companies, many of which entered **Series A** rounds.

1 PRE-SEED

Mission Control	\$2.0m
-----------------	--------

2 SEED

Anch AI	\$2.1m
Checkstep	\$5.0m
Datatron	\$12m
Fairgen	\$2.5m
Pendella	\$5.2m
Sapia	\$7.0m
Sarus	-----
Trustero	\$8.0m

3 SERIES A

Aporia	\$25m
Bodyguard AI	\$11m
CausaLens	\$53m
Credo AI	\$13m
Diversio	\$6.3m
FairPlay AI	\$10m
Mathison	\$25m
Netacea	\$11m
Reejig	\$15m
Synthesis AI	\$17m
XOPA AI	\$4.2m

4 SERIES B

Cohere	\$122m
Datagen	\$50m
Mostly AI	\$25m
Spectrum Labs	\$32m
TruEra	\$25m

5 SERIES C-E

Flock Safety	\$150m
MDCClone	\$63m
Pave	\$100m

6 UNDISCLOSED

Ambient AI	\$52m
Equalture	\$2.0m
Eskalera	-----
Inrupt	-----
Mindtech Global	\$3.7m
Pipeline Equity	-----
Prifina	\$0.3m
QuadFi	\$99m
Textio	\$1.0m
Troj AI	\$2.3m
vAlsual	-----

Some of the most well-known venture and investment firms have made several forays into the Ethical AI space. **Tiger Global** leads with five separate investments.

TIGERGLOBAL

 **aporia**  flock safety
 crosschq  co:here
 ROBUST INTELLIGENCE

 **Lightspeed**

 snorkel  eightfold.ai
CALYPSO AI

GGVCAPITAL

TONIC  snorkel

 **Y** Combinator

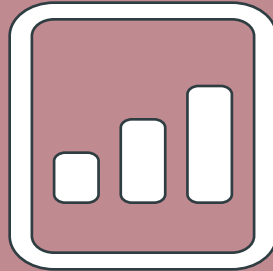
PAVE  HUMANLY
ZUMOLABS

 **Index**
Ventures

PAVE  Arthur
 BrightHire

 greylock**partners.**

truera  gretel



Category Trends

Data for AI
AI Audits, Governance, Risk, and Compliance
ModelOps, Monitoring, and Observability
Targeted Solutions and Technologies

Five EAIDB Categories



Data for AI

- Data Privacy
 - anonymization / privatization
 - safe data transfer
 - data permissioning
- Synthetic Data
- Data Bias Detection
- Ethical Data Sourcing
 - surveying
 - labeling



AI Audits & GRC

- Consulting Firms
 - ethical data strategy
 - ethical AI advising
- Governance & Compliance
 - automated platforms
 - business to data KPI conversion
- Risk
 - AI risk management
 - risk quantification



MLOps & Monitoring

- In-Training Bias Monitoring
- Explainability
- Automated Model Diagnostics



Targeted Solutions

- FinTech, HireTech, InsurTech, HealthTech
- Proprietary Technology
 - ex. toxic content detector algorithms
 - ex. ethical facial recognition algorithms



Open-Source

- Any fully open-source framework that fits into any of the above categories

Synthetic Data

✓ Text and tabular data is generalizing.

The nature of text / tabular data is that it is not tied to a particular domain (i.e., it is domain-agnostic). Generating text about vehicles, for example, is not significantly more complicated than generating text about shopping. Text, numbers, and distributions are easier to re-create because they have less dimensional complexity when compared to, say, vision. Companies like Fairgen [2], Clearbox AI [3], and others are working on generalizing their algorithms to more horizontals. Fairgen, for instance, is working on branching out to include text generation.

✓ Vision data is narrowing.

On the other hand, vision data is highly domain-dependent. Generating images of houses, for example, requires domain expertise (including multi disciplinary contributions from architects, 3D renderers, etc.) that is very different from that required by human face generation. As a result, companies in this space focus on a specific vertical (e.g., industrial vision sets, facial datasets, etc.). Naturally, the companies that are able to harness the largest number of verticals stand to benefit. In the words of Ofir Zuk from Datagen [4], "it's hard to reach critical mass by crawling through verticals. Expanding across domains is a better way to scale."

⊗ Synthetic data markets are affected by lack of uptake.

Synthetic data is a boon to companies worried about data privacy or data bias. It has even been shown to improve the overall performance of machine learning algorithms. However, there are certain highly-regulated industries that are not quite ready to accept the benefits of synthetic data. In Europe, banks and telecom companies are quick to use synthetic approaches, but healthcare companies are showing hesitation. Being a safe-critical field with very strict privacy requirements, the healthcare sector has no consensus on the use of sensitive data. Coupled with Europe's lagging AI maturity (relative to the USA), the lack of uptake and awareness of synthetic data is currently a significant barrier.

Data Bias, Sourcing, and Observability

✓ The need for **good seed data** is (and will always be) a priority.

In the data for AI industry, high-quality seed datasets are still a top priority because synthetic sets can only be generated from a seed set. The process of curating a dataset today is machine-assisted (through providers like Snorkel [5]), but is increasingly involving humans. Companies like co:census [6] attempt to provide another option for data gathering through the use of surveys, public comments, audio transcriptions, and social media. Companies source humans to validate their data with edge cases and adversarial examples.

✗ Good, **cross-cultural** data collection is not yet prioritized.

Data is collected differently depending on where in the world it is sourced from. Cross-cultural influences must be considered when data is collected. Unfortunately, as is the case with most subfields in AI, most companies handling data prior to ML applications are based in North America and Europe, with little to no African or South American representation. For nuanced datasets, such as facial sets in vision, cultural differences are key. Gaining diverse perspectives and making sure data is represented well in a cultural way is something no ML-assisted curation method is good at as of now.

✗ Lack of **context-consciousness** for data bias detection.

Data is formed from external methodologies and is ingrained with the circumstances of its collection. Checking for one type of data bias is easy—underrepresented protected classes, for example—but building a holistic platform that looks in the data and corrects multiple types of biases is very hard. The largest piece of the puzzle is context. Understanding the context of where the data came from and what other lurking biases may be present is a ridiculously hard problem to automate because it varies on a case-by-case basis. A platform must know, for example, that crime statistics in the United States are skewed severely against racial minorities in order to understand the extent of bias in criminal justice data.

MLOps, Monitoring, and Observability



TREND



CHALLENGE



Many MLOps-specific companies are **expanding horizontally**.

Due to market conditions, namely low demand and high fragmentation, many of EAIDB's MLOps startups are pivoting horizontally to encompass more of the machine learning pipeline. These companies may start with pure MLOps, but then add governance or risk analysis capabilities on top. This horizontal expansion allows businesses to increase their total addressable market (TAM). MLOps companies are generally the most technically challenging; the high barriers to entry mean that they are best positioned for a horizontal takeover of the responsible machine learning pipeline.



Most early-stage MLOps companies are **focusing on highly regulated industries**.

As of now, there are only a few sectors that directly deal with human society in such a way that fairness, transparency, and responsibility come to the forefront. These areas also tend to already have strict regulations that drive demand for compliance and risk analysis tools. In the United States and Europe, finance, healthcare, and insurance are some of these sectors. Most startups in the MLOps category begin their journeys here, but this may evolve as new regulations are passed. As new technologies like Web 3.0 or metaverses continue to meld with human society, more sectors will emerge and will need regulation.



Highly fragmented space with **fairly low demand**.

Data is formed from external methodologies and is ingrained with the circumstances of its collection. Checking for one type of data bias is easy—underrepresented protected classes, for example—but building a holistic platform that looks in the data and corrects multiple types of biases is very hard. The largest piece of the puzzle is context. Understanding the context of where the data came from and what other lurking biases may be present is a ridiculously hard problem to automate because it varies on a case-by-case basis. A platform must know, for example, that crime statistics in the United States are skewed severely against racial minorities in order to understand the extent of bias in criminal justice data.

AI Audits, Governance, Risk, and Compliance

✓ Automated GRC platforms are gaining traction, but ethics-specific professional service firms **are not going anywhere**.

As with any other industry, there is a strong trend towards automating governance, risk, and compliance. The rise in automated platforms broadens the number of available tools to operationalize governance and risk, but companies often turn to consulting firms and professional service firms to figure out how exactly to use these tools for their specific problem. Both automated and traditional solutions will continue to grow in tandem, especially in the EU where companies are still ramping on AI usage.

✓ As AI auditing becomes more legitimate, organizations may be required to perform **third-party AI audits as a compliance rule**.

There is a growing movement of individual AI auditors that are gaining legitimacy and becoming certified. ForHumanity [7], a nonprofit organization focusing on education and training for these auditors, is working with the Information Commissioner's Office (ICO, UK) [8] to formalize third-party AI auditing as a compliance requirement for companies. This will be a driving force for many of the companies in EAIDB that handle AI auditing (primarily consulting firms), but this change also benefits the entire industry by increasing awareness and focus on the issue.

⊗ **Competition is increasing** as startups expand into GRC.

MLOps companies are using horizontal expansion to assuage demand lulls. Opening a product to include AI GRC is a good way to increase the Total Addressable Market (TAM) of the business. Unfortunately for existing automated GRC solutions, this means there is an influx of competition that increasingly fragments the market. Demand may further fracture, but the incumbent firms with established presence in particular regions (i.e. 2021 AI [9] in the Nordics, Mind Foundry [10] in the UK, etc.) have the first mover advantages of trustworthiness and track record, both of which are critical in this industry.

Targeted Solutions and Technologies



TREND



CHALLENGE



The Hiretech space continues to see **new entrants**.

The hiretech boom, which began in 2015 or so, continues to fuel this category even today as new firms with better AI solutions enter the ecosystem. However, there are estimates that the next few years will see some level of defragmentation via acquisitions and exits. Some larger incumbents, like Pave [11] have already begun to consolidate - Pave acquired Advanced-HR [12] in June earlier this year.



Startups tackling unfairness in highly regulated industries enjoy high barriers to entry and, therefore, **little competition**.

At the core of the issue is the idea that regulated industries have extremely high inertia - it takes a Herculean effort to enact any change at all. Regulators are slow to move and the existence of unfairness or non-ethical case studies must be proven and accepted as fact by the entire industry before the runway can be cleared and action can be taken. Only recently has unfairness in finance been revealed. Startups like FairPlay AI [13], Evispot [14], QuadFi [15], and Stratyfy [16] were founded within the last five years. The evidence is mounting in healthcare as well, though there are few (if any) companies focused specifically on fairness in healthcare.



Regulators in industries like healthcare, insurance, and lending may have **conflicting or outdated definitions of fairness**.

In lending, for example, most regulators are familiar with a metric called the Adverse Impact Ratio (AIR). This measure compares the rates of loan issuance between two different groups. However, it is well-known in technical circles that this is an extremely crude metric that does not capture the full extent of unfairness in lending. The lack of shared language between technical specialists and regulators renders it nearly impossible to demonstrate that there is unfairness in a system (especially because institutions will make sure that they pass the AIR test and neglect stronger evidence of unfairness).

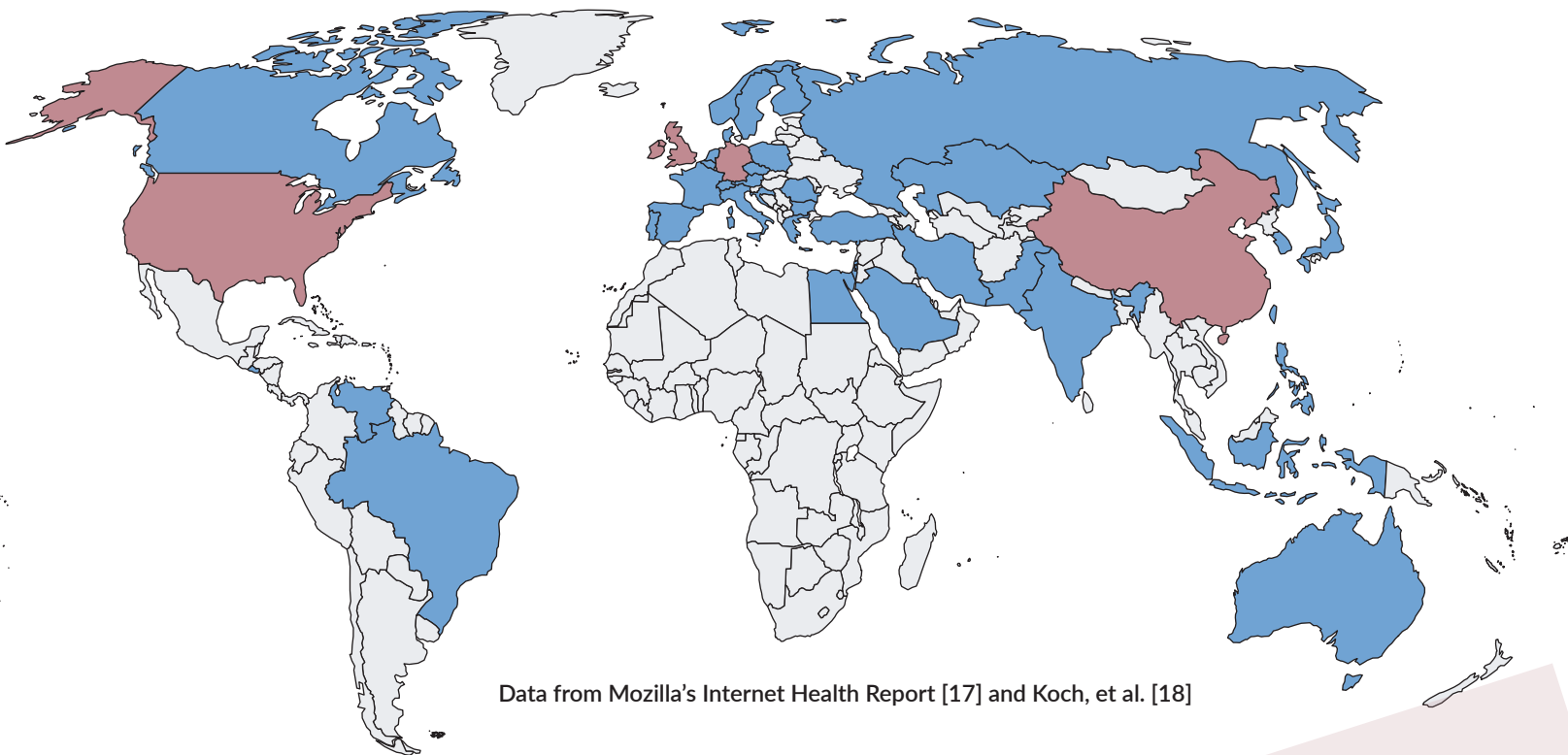


Barriers to Ethical AI

Lack of Cultural Diversity
Demand for Responsible AI

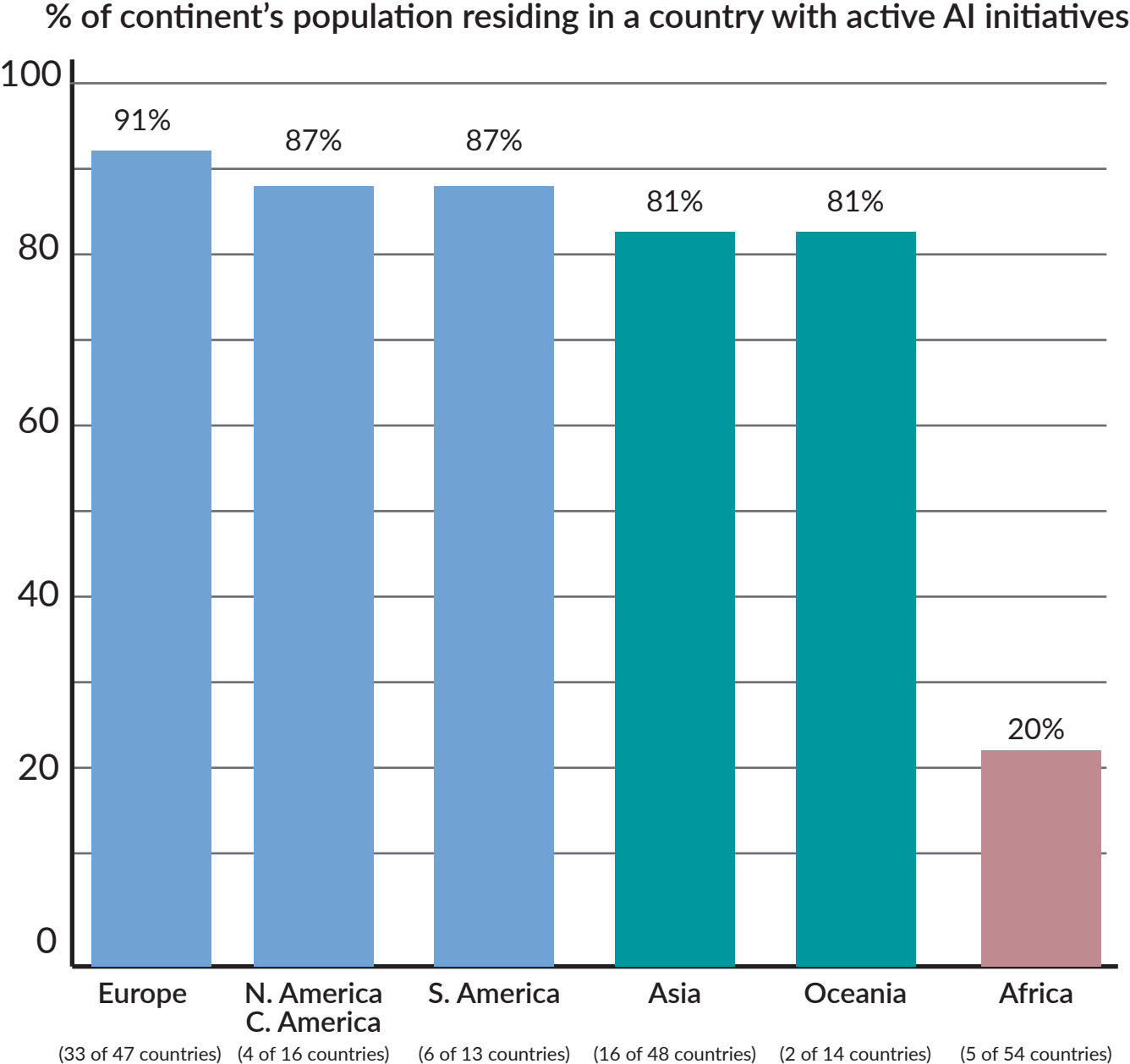
Datasets sourced from countries who are already at the top in AI are the only ones that are repeatedly used for benchmarking purposes. Because these nations tend to be Western countries, the datasets used by the AI community to build and compare models are almost **completely devoid of non-Western cultural perspective.**

Benchmarking Dataset Usage, Global Distribution



- Datasets sourced from this country have been used 1,000+ times.
- Datasets sourced from this country have been used at least once.
- Datasets sourced from this country have never been used.

The distribution of countries with AI initiatives is a good proxy for global participation in AI discussion and regulation. A closer look reveals that **African perspective is essentially missing.**



The global discussion is also lacking input from less populous cultures within each of these continents. Most of **Africa, Central America, Middle East,** and the **Pacific Islands** are absent as well. In general, this means that current and future AI policy will be designed with only the majority in mind (i.e. policy to combat bias itself becomes biased). Inclusivity on the global stage from diverse cultures is critical in the sphere of ethical AI.

Global Demand for Ethical AI

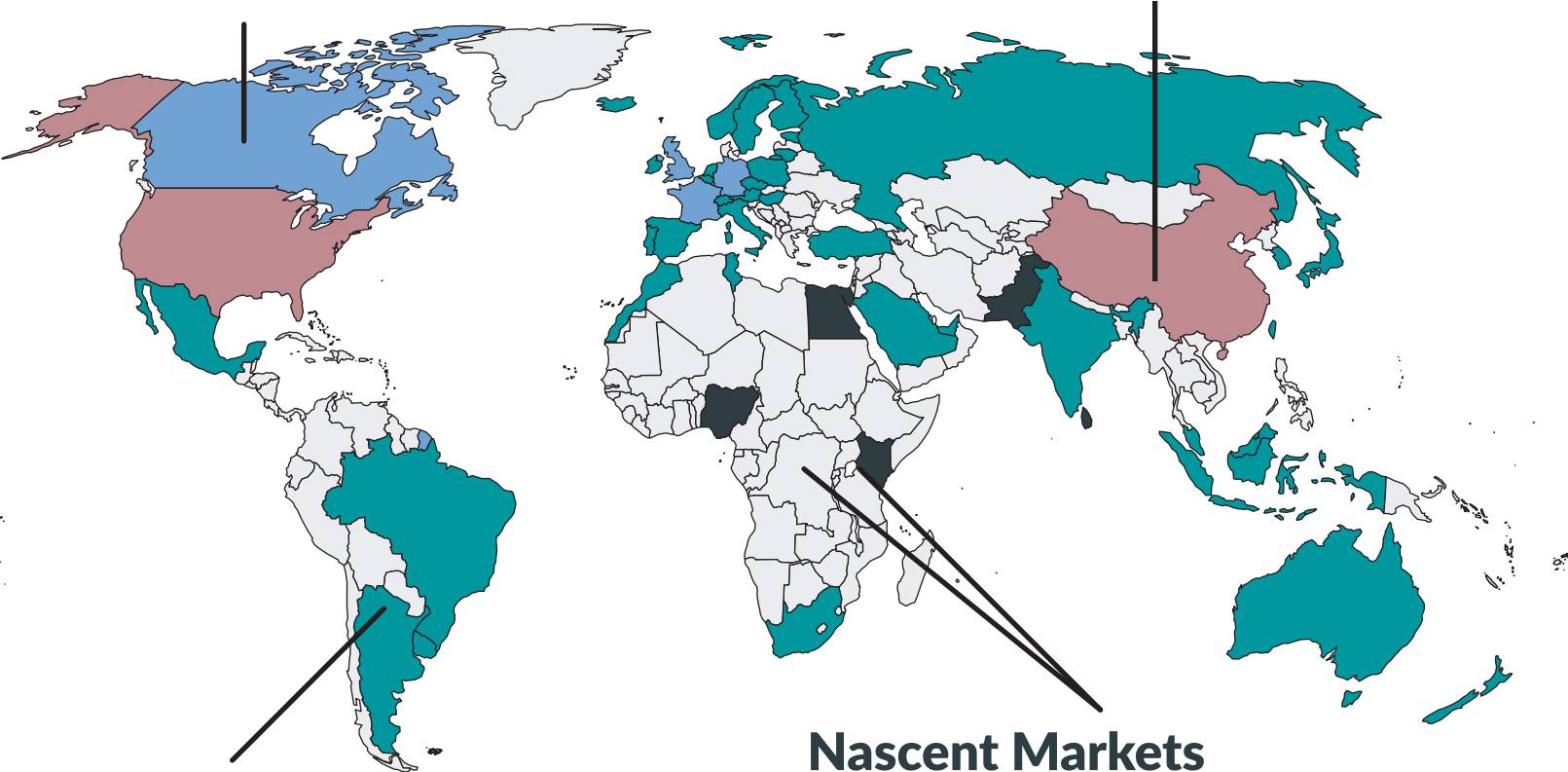
Despite growing in popularity across the globe, there is still very limited demand for ethical AI.

Established Markets

Canada and the European Union lead the world in AI regulation and policy. In these markets, regulatory pressure exists and will continue to raise the Ethical AI market for years to come. However, the AI is so much less mature in these areas than in the World AI Powers that clients are still not aware of what their main pain points are. This severely limits the demand for Ethical AI solutions in the near-term.

World AI Powers

In the United States and China, AI markets are mature but policy severely lags other regions like Europe or Canada. There is a steadily increasing demand for ethical AI, but the market is severely limited by lack of regulatory pressure and general unwillingness to invest in tangential technologies that do not directly further a company's business model.



Rising Markets

Countries like Argentina, Australia, India, and Mexico are just now ramping up their AI usage. Ethics is not yet in the minds of most organizations because their AI markets are immature, but most regions in this category do have national AI strategies. Even so, demand for ethical AI is extremely low.

Nascent Markets

Little to No Market

Regions like Kenya, Egypt, Pakistan, and Nigeria are exploring AI and the advantages it can bring. Most nations in Africa, Central America, the Pacific Islands, and the Middle East simply have no AI market whatsoever. Understandably, there is little to no demand for ethical AI here.

Ongoing EAIDB Initiatives

MAIEI

Category deep-dives with the [Montreal AI Ethics Institute](#) [22].

An article series exploring each of EAIDB's categories in rich detail with additional color provided by founders. The most recently published article covered the "Data for AI" category with commentary from Fairgen [2], Clearbox AI [3], and Datagen [4].

[Read](#)



Nordic Innovation

Mapping the Nordic Ethical AI ecosystem with [Nordic Innovation](#) [23].

Nordic Innovation is a division of the Nordic Council of Ministers and promotes entrepreneurship, innovation, cooperation, and competitiveness in the five Nordic countries. EAIDB is working to map the ethical AI ecosystem (including startups and institutions).

Interested in collaborating with us? Reach out on <https://eaidb.org>.

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

[1] Ethical AI Database (EAIDB)	https://eaidb.org
[2] Fairgen	https://fairgen.ai
[3] Clearbox AI	https://clearbox.ai
[4] Datagen	https://datagen.tech
[5] Snorkel	https://snorkel.ai
[6] co:census	https://cocensus.io
[7] ForHumanity	https://forhumanity.center
[8] ICO	https://ico.org.uk
[9] 2021 AI	https://2021.ai
[10] Mind Foundry	https://mindfoundry.ai
[11] Pave HR	https://pave.com
[12] Advanced-HR	https://advanced-hr.com
[13] FairPlay	https://fairplay.ai
[14] Evispot	https://evispot.com
[15] QuadFi	https://quadfi.com
[16] Stratyfy	https://stratyfy.com
[17] Internet Health Report	https://2022.internethealthreport.org
[18] Koch, et al. (2021)	https://arxiv.org/pdf/2112.01716.pdf
[19] HolonIQ	https://holoniq.com
[20] OECD AI	https://oecd.ai
[21] Tortoise Media	https://tortoisemedia.com
[22] MAIEI	https://montrealetics.ai
[23] Nordic Innovation	https://nordicinnovation.org