Machine Learning Practices Outside Big Tech:

How Resource Constraints Challenge Responsible Development

Aspen Hopkins* & Serena Booth*
MIT CSAIL

AI, Ethics, & Society (AIES) 2021

October 23, 2020

Anger Builds Over Big Tech's Big Data Abuses

Alex Woodie

The surveillance economy c
coming under greater scrutir
the Department of Justice's
Google this week and films l
which was released last mor
Internet clearly are angry, bu
action? That remains to be s

October 23, 2020

Anger Builds Over Big Tech's Big Data Abuses

Alex Woodie

The surveillance economy c
coming under greater scruti
the Department of Justice's
Google this week and film
which was released last
Internet clearly are angr
action? That remains to be s

Largest fine under GDPR levied against Goog

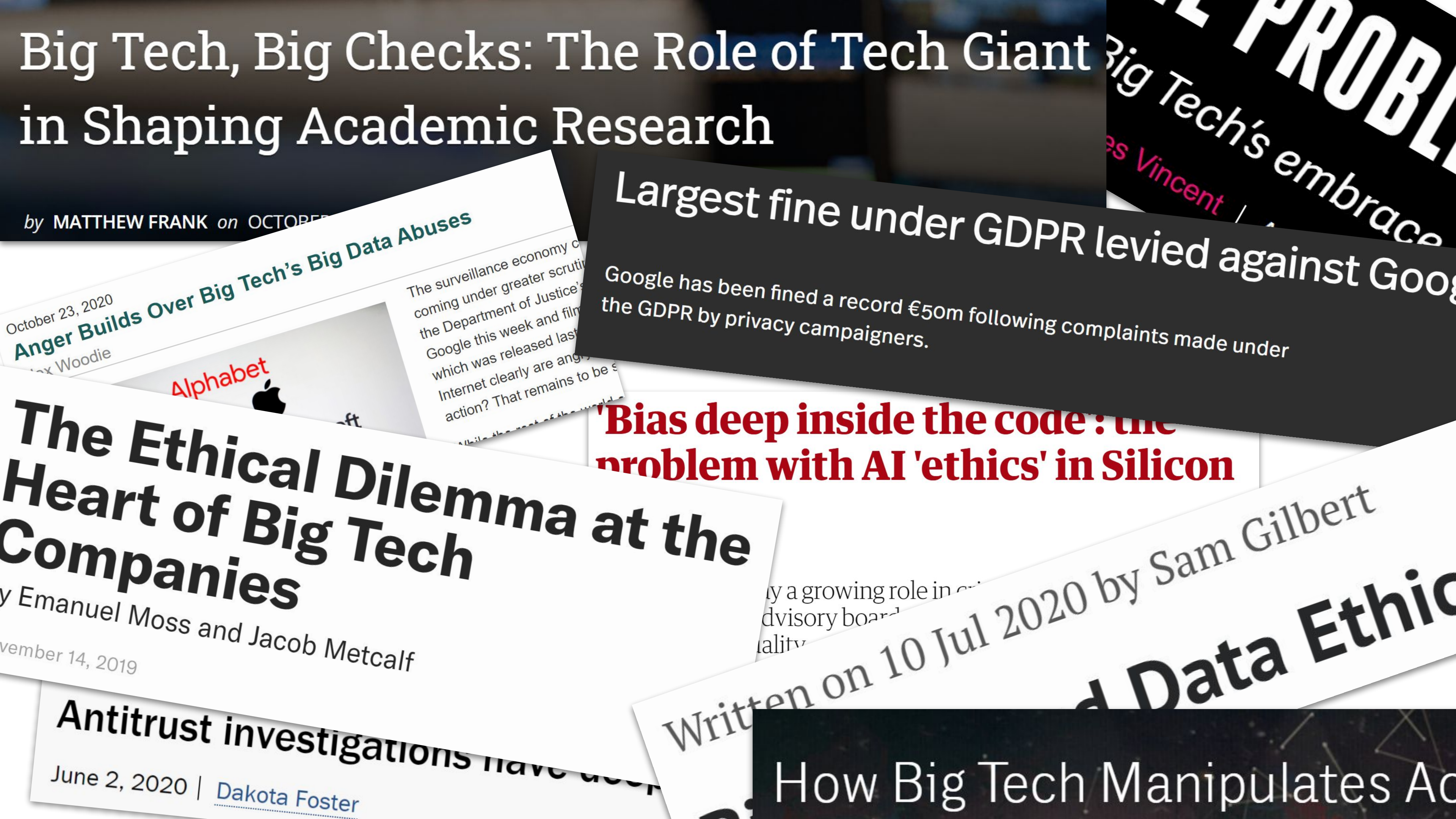
Google has been fined a record €50m following complaints made under the GDPR by privacy campaigners.

'Bias deep inside the code': the problem with AI 'ethics' in Silicon Valley

As algorithms play a growing role in criminal justice, education and more, tech advisory boards and academic programs mirror real-world inequality

Antitrust investigations have deep implications for AI and nati

June 2, 2020 | Dakota Foster



Big Tech, Big Checks: The Role of Tech Giant in Shaping Academic Research
by MATTHEW FRANK on OCTOBE
Big Tech's embrace
Vincent
Largest fine under GDPR levied against Goog
Google has been fined a record €50m following complaints made under the GDPR by privacy campaigners.
October 23, 2020
Anger Builds Over Big Tech's Big Data Abuses
Alex Woodie
The surveillance economy
coming under greater scruti
the Department of Justice's
Google this week and film
which was released last
Internet clearly are angr
action? That remains to be
Alphabet
'Bias deep inside the code': the problem with AI 'ethics' in Silicon
The Ethical Dilemma at the Heart of Big Tech Companies
Emanuel Moss and Jacob Metcalf
vember 14, 2019
y a growing role in
dvisory board
Written on 10 Jul 2020 by Sam Gilbert
Data Ethic
Antitrust investigations
June 2, 2020 | Dakota Foster
How Big Tech Manipulates Ac

But what about everyone else?

“The long tail of ML deployment is where oversight is likely to be least, yet the potential for harm remains high”

the big questions

the big questions

How do development practices of “long tail” organizations compare to Big Tech?

the big questions

How do development practices of “long tail” organizations compare to Big Tech?

How can we align research to (better) encompass these “long tail” practitioners?

a few barriers

a few barriers

a few barriers

Fewer resources

a few barriers

Fewer resources

Added pressure from increased *existential risk*

[Svenja, Loch, and Dong 2009]

a few barriers

Fewer resources

Added pressure from increased *existential risk*
[Svenja, Loch, and Dong 2009]

Less AI/ML experience

a few barriers

Fewer resources

Added pressure from increased *existential risk*
[Svenja, Loch, and Dong 2009]

Less AI/ML experience

Difficulties Hiring

Type	Company Description	Interviewee Title	Resources
Publicly Listed	Shopping/recommendations	Data Engineer	$$$$
Startup	Shopping/recommendations	VP of Product	$$
Startup	Shopping/recommendations	VP of Strategy	$$
Publicly Listed	Pet care (diagnostics)	Senior Data Scientist	$$$$
Startup	Healthcare (diagnostics)	Chief Operating Officer	$
Startup	Fitness	Chief Technology Officer	$$$
Startup	Real estate	Chief Technology Officer	$$
Small Company	Real estate	Head Of Analytics	$$
Startup	Real estate	Senior Product Manager	$$
Startup	ML consulting and tools	Chief Technology Officer	$$
Startup	ML consulting and tools	Chief Executive Officer	$
Startup	Data automation	Board Member/Investor	$
Startup	Pet care	Director of Engineering	$
Public Sector	Municipality	Asst. Director of Data Analytics	$
Venture Capital	Investment	Startup/ML Investor	-
Startup	Language learning	Chief Technology Officer	$
Startup	Language learning	Chief Technology Officer	$

Interviews with diverse orgs

(semi-structured) interview questions

- How do you evaluate your models or data?

(semi-structured) interview questions

- How do you evaluate your models or data?
- What do updates and changes to your models or data look like?

(semi-structured) interview questions

- How do you evaluate your models or data?
- What do updates and changes to your models or data look like?
- How has GDPR or other data-related legislation affected you?

(semi-structured) interview questions

- How do you evaluate your models or data?
- What do updates and changes to your models or data look like?
- How has GDPR or other data-related legislation affected you?
- ...

thematic analysis

- Focused on ML or on data services
    - Team does the ML and provides the results to other teams and vets
- What do your models do and what is a challenge?
    - Cleaning data has been the biggest challenge. Our early ML was focused on what products/services were sold on a given day
    - This means lines on an invoice
    - First models were on standardizing invoice line items from the practices
    - Built a multilabel hierarchical classification service
        - Still running
    - Latest is models for predicting disease risk using lab results and other clinical results
    - Some things on NLP but just beginning--identifying errors in transcribed voice texts (radiologists dictate to drag in software, but it makes mistakes because its not trained on radiology terminology. So they're training to identify post-processing potential errors)
- Concern about radiologists trusting ML models over expertise / too much
    - From their perspective, disease predictive models are first place that comes into play. Working with designers to think about how it might not happening
    - Explainability is the to build more trust, that's the hump a lot of work has focused on
    - However we thinking about how we present this information in a non-definitive way (not this animal will have hypothyroidism, as an example)
    - To validate model, they do standard dataset split but then also have internal review to check the model results and labels
        - This matters less for line items

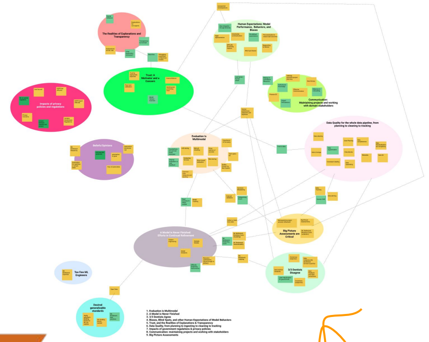
inherent risk

Aspen Hopkins 7:52 PM Jan 1
Opinion: Auditing is important for high risk use cases only

Aspen Hopkins 7:53 PM Jan 1
Opinion: young companies don't think about GDPR/ needs for explanations

Aspen Hopkins 7:53 PM Jan 1
Opinion: only big companies are effected by GDPR

Aspen Hopkins 7:54 PM Jan 1
Opinion: big teams needed for useful explanations



Algorithm			
best-practices	best practices in reflection		
Challenge	Bias mitigation through diverse users	A sufficiently diverse userbase will protect against bias.	*
Challenge	big data operations on server is expensive	Sending big data to a server is expensive	*
Challenge	Privacy concerns with user data on server	Sending user data to a server compromises privacy	
Challenge	customer concerns of expertise/trust	Building sufficient trust.	*
Challenge	data cleaning	Cleaning data.	
Challenge	data comprehensiveness	Missing records, record consistency and completeness.	*
Challenge	inconsistant labeling	Consistent data entry and tagging systems	
Challenge	inconsistant labeling	Consistent labeling	
		Need consensus, even for subjective	

Explanations for non-experts	Documentation
Data Planning	Documentation
Project Permanence	Documentation
External validators	Performance
Importance of Accuracy	Performance
Need accuracy to reflect context	Performance
Model failures - trust lost	Performance
Multiple Models	Performance
Model iterations	Performance
Personalization to reduce bad outcomes	Performance
Privacy Concerns	Privacy
tension between ubiquity and privacy	Privacy
GDPR hasn't affected	Privacy
GDPR has affected	Privacy
Trust by appropriate expertise	Trust
Poor trust calibration	Trust
Black Box	Trust
Trust in data	Trust

6 themes

1. Expectations vs Feasibility
2. Black Boxes, Explanations, & Overconfidence
3. A Model is Never Finished
4. Assessing, Preventing, & Mitigating Bias
5. Communication & Collaboration
6. Privacy vs Growth

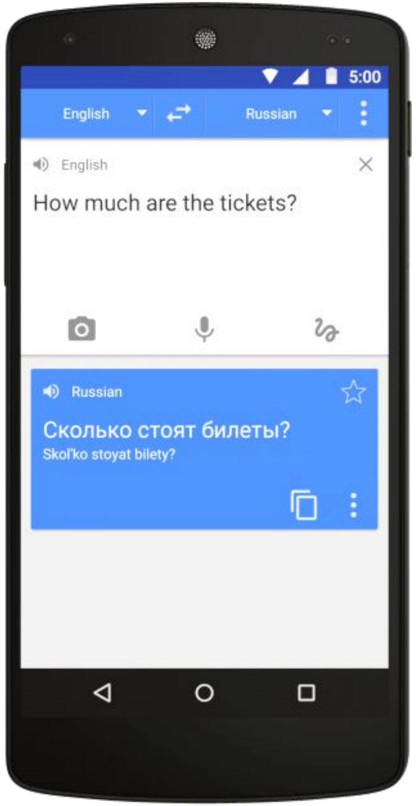
1. Expectations vs Feasibility

1.1 User Expectations

Big Tech has a **participation monopoly**

"Users expect [us] to be equal or better to Google translate."

Increased existential risk = **increased pressure to act**

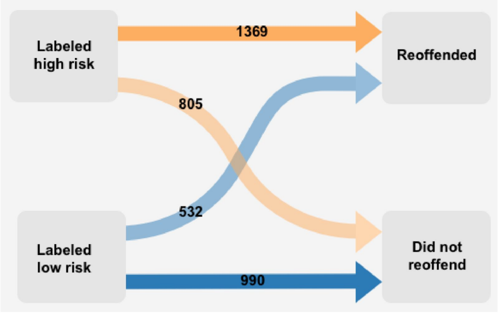


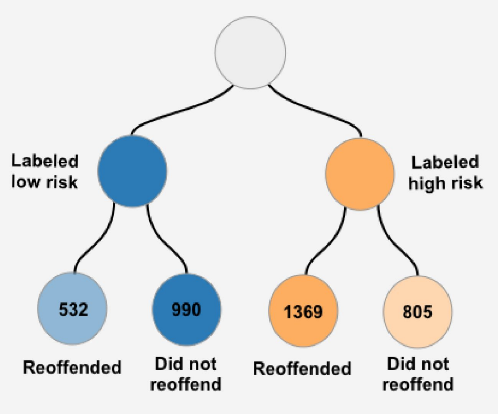
2. Black Boxes, Explanations, & Overconfidence

2.4 Mitigating Overconfidence

LIME and feature importance explanations were “unhelpful”--“feature importance sucks”

Seeking to present information in a “non-definitive” way as an alternative to formal explanations

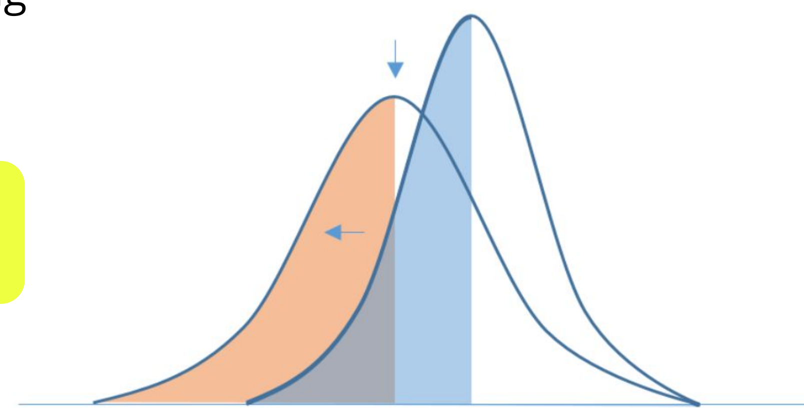




[Shen et al., 2020]

3. A Model is Never Finished

4.5.1 Data Quality: Planning, Ingesting, & Cleaning
4.5.3 Model and Data Versioning



1. “Lack of best practices in training”

2. Trusted data minimizes costs

4. Assessing, Preventing, & Mitigating Bias

4.1 Bias Mitigation Through Diversity or Personalization

Rather than mitigating bias post/in-training, all interviewees focused on *data,* not models

…But have few developed standards for data collection and quality

4. Assessing, Preventing, & Mitigating Bias

4.2 Assessing Blind Spots

A troubling trend of *deferred responsibility,*

Complacency for apparently low risk: “might mean a $XXX medical procedure instead of an $XX medical procedure”

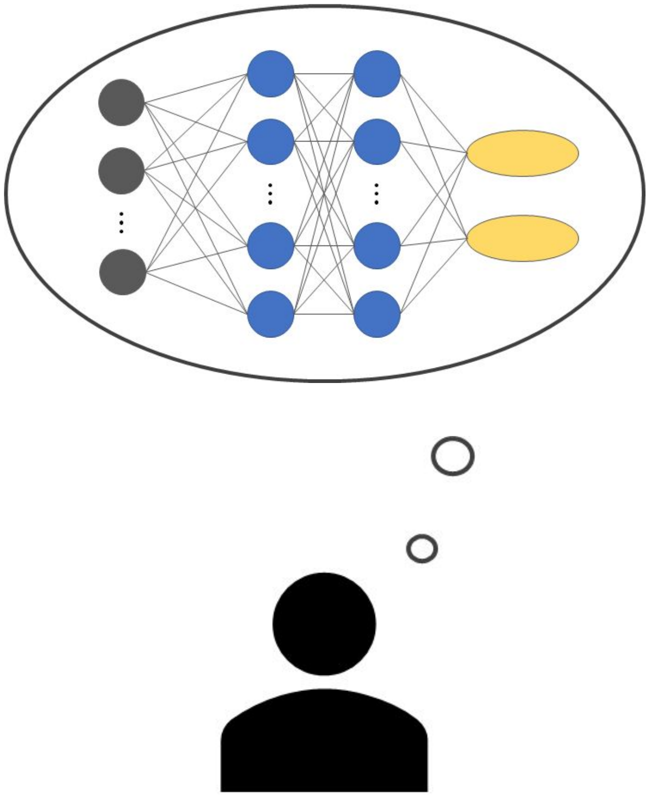
6. Privacy vs Growth

4.6.1 Government Regulation

“GDPR doesn’t affect us”

all interviewees expressed this sentiment

- Companies aren’t prepared for right to explanation / transparency
- They retain trained models
- Deletion requests are considered a large burden, though desired



What did we find?

While orgs outside big tech face many shared challenges to responsible development, difficultes are exacerbated by resource constraints and…

While orgs outside big tech face many shared challenges to responsible development, difficultes are exacerbated by resource constraints and…

- big tech’s monopoly on AI/ML participation

While orgs outside big tech face many shared challenges to responsible development, difficultes are exacerbated by resource constraints and…

- big tech’s monopoly on AI/ML participation
- lacking tooling/guidelines for smaller-scale dev

While orgs outside big tech face many shared challenges to responsible development, difficultes are exacerbated by resource constraints and…

- big tech’s monopoly on AI/ML participation
- lacking tooling/guidelines for smaller-scale dev
- reduced concern for GDPR requirements

While orgs outside big tech face many shared challenges to responsible development, difficultes are exacerbated by resource constraints and…

- big tech’s monopoly on AI/ML participation
- lacking tooling/guidelines for smaller-scale dev
- reduced concern for GDPR requirements
- increased sense of *deferred responsibility*

Thanks!

Twitter:

@aspenkhopkins
@SerenaLBooth

Email:

dataspen@mit.edu
sbooth@mit.edu

Machine Learning Practices Outside Big Tech: How Resource Constraints Challenge Responsible Development

Aspen Hopkins*
Massachusetts Institute of Technology
Cambridge, USA
dataspen@mit.edu

Serena Booth*
Massachusetts Institute of Technology
Cambridge, USA
sbooth@mit.edu

ABSTRACT

Practitioners from diverse occupations and backgrounds are increasingly using machine learning (ML) methods. Nonetheless, studies on ML Practitioners typically draw populations from Big Tech and academia, as researchers have easier access to these communities. Through this selection bias, past research often excludes the broader, lesser-resourced ML community—for example, practitioners working at startups, at non-tech companies, and in the public sector. These practitioners share many of the same ML development difficulties and ethical conundrums as their Big Tech counterparts; however, their experiences are subject to additional under-studied challenges stemming from deploying ML with limited resources, increased existential risk, and absent access to in-house research teams. We contribute a qualitative analysis of 17 interviews with stakeholders from organizations which are less represented in prior studies. We uncover a number of tensions which are introduced or exacerbated by these organizations' resource constraints—tensions between privacy and ubiquity, resource management and performance optimization, and access and monopolization. We argue that increased academic focus on these lesser-resourced practitioners can facilitate a more holistic understanding of ML limitations, and so is useful for prescribing a research agenda to facilitate responsible ML development for all practitioners.

CCS CONCEPTS

• **Social and professional topics → Socio-technical systems**; **Computing organizations**; **Codes of ethics**.

ACM Reference Format:
Aspen Hopkins and Serena Booth. 2021. Machine Learning Practices Outside Big Tech: How Resource Constraints Challenge Responsible Development. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society (AIES '21), May 19–21, 2021, Virtual Event, USA.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3461702.3462527

1 INTRODUCTION

ML practitioners are increasingly composed of people from diverse occupations and backgrounds. Yet, in past research analyzing ML practice, the vast majority of studies draw participants from Big Tech companies or academia [1, 5, 22, 24, 28, 29, 29, 30, 36, 37, 41, 45, 58, 60], with few exceptions [9, 25, 42]. However, wealthy Big Tech and academic communities offer privileges and perspectives that are *not* universally representative. For example, Simonite [50] chronicled how a Google and Carnegie Mellon University project collected 300 million labels and used fifty GPUs for two months—a scale of development which is increasingly the norm, yet is untenable for less resourced or experienced organizations. This leads to the question: how well do past studies of Big Tech and academic practitioners encompass the needs of other data and ML workers?

Pereira et al. [42] observed that the diversity of data science teams' composition, goals, and processes remains understudied—particularly for practitioners outside of Big Tech. We note this is certainly not the only understudied component of data and ML work outside of Big Tech and academia, and ask: what are the problems smaller companies, organizations, and agencies face? What are their practices? How can we, the AI research community, ensure that the work we do is targeted not just to benefit well-resourced organizations but also those with limited fiscal resources and increased *existential risk*, where any given decision may carry the added risk of not making payroll [51]? These questions are particularly consequential to future work encouraging ethical and fair practices [12], as these organizations often find applying current best practices in responsible AI development to be too costly.

We conducted 17 interviews with practitioners working outside of Big Tech and academia, asking questions about current practices, fairness, and risk mitigation in ML development. We analyzed these semi-structured interviews using thematic analysis, uncovering six themes and numerous insights about these practitioners' beliefs and behaviors. We explore tensions between privacy and ubiquity, resource management and performance optimization, and access and monopolization. We focus on the impacts (or lack thereof) of GDPR and privacy legislation, the limited usefulness of model explanations, the trend of deferring responsibility to downstream users and domain experts, and Big Tech's monopolization of access. These tensions reflect organizations' underlying and competing concerns of growth and cost, with frequent and complex trade-offs.

While our findings often overlap with those of past practitioner studies, we find that resource constraints introduce additional challenges to developing and testing fair and robust ML models. Fur-

*Both authors contributed equally to this research.

Learn more in our paper!