Generative AI + Law
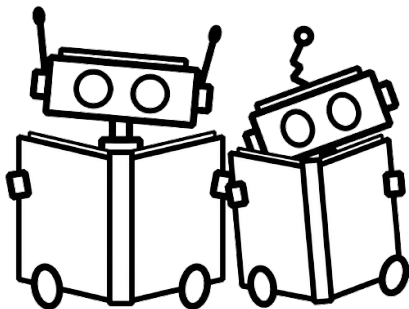
ICML: July 29, 2023

Goals

Precision
Common language
Shared research directions

Time	Speaker
9:15 - 9:45 AM	
9:45 - 10:05 AM	
10:05 - 10:40 AM	
10:40 - 11:00 AM	IP
11:00 - 12:00 PM	
12:00 - 1:30 PM	
1:30 - 1:45 PM	
1:45 - 2:15 PM	
2:15 - 3:00 PM	Spotlights + Posters
3:00 - 3:30 PM	
3:30 - 3:45 PM	
3:45 - 4:00 PM	Privacy
4:00 - 5:00 PM	

Generative AI + Law

GenLaw '23

Time	Speaker
9:15 - 9:45 AM	Pamela Samuelson, "Some Nonobvious Observation About Copyright's Scope For Generative AI Developers"
9:45 - 10:05 AM	Mark Lemley, "Is Training AI Copyright Infringement?"
10:05 - 10:40 AM	**Coffee**
10:40 - 11:00 AM	Miles Brundage, "Where and when does the law fit into AI development and deployment?"
11:00 - 12:00 PM	*Panel on Intellectual Property*: Pamela Samuelson, Mark Lemley, Luis Villa, Katherine Lee (Moderated by Jack Balkin and A. Feder Cooper)
12:00 - 1:30 PM	**Lunch**
1:30 - 1:45 PM	Jack Balkin, "Artificial Intelligence and the First Amendment"
1:45 - 2:15 PM	**Spotlights**
2:15 - 3:00 PM	**Posters**
3:00 - 3:30 PM	**Coffee**
3:30 - 3:45 PM	Nicholas Carlini, "A Brief Introduction to Machine Learning & Memorization"
3:45 - 4:00 PM	Gautam Kamath, "What does Differential Privacy have to do with Copyright?"
4:00 - 5:00 PM	*Panel on Privacy*: Kristen Vaccaro, Nicholas Carlini, Miles Brundage, Gautam Kamath, and Jack Balkin (Moderated by Katherine Lee and Deep Ganguli)

Generative AI + Law

GenLaw '23

Speakers, Panelists & Moderators

Pamela Samuelson

Distinguished Professor of Law and Information

University of California, Berkeley

Mark Lemley

Professor of Law

Stanford Law School

Nicholas Carlini

Research Scientist

Google Brain

Gautam Kamath

Assistant Professor

University of Waterloo

Kristen Vaccaro

Assistant Professor

University of California, San Diego

Luis Villa

Co-founder and General Counsel

Tidelift

Miles Brundage

Head of Policy Research

OpenAI

Jack M. Balkin

Professor

Yale Law School

GenLaw '23

What are we talkin' about?

Dataset creation

Model training

Model serving

Dataset creation

Model training

Model serving

Training **dataset creation** involves a lot of *choices*

What type of data?
How much data?

What type of data?
How much data?
Web data?
Which web data?

What type of data?
How much data?
Web data?
Which web data?
"Good" data?

What type of data?

How much data?

Web data?

Which web data?

"Good" data?

Copyrighted data?

Private data?

Training **dataset creation** involves a lot of *choices*

The Devil is In the Training Data

Chapter 1

AUTHORS

Katherine Lee

Daphne Ippolito

A. Feder Cooper

The process of training contemporary generative models requires vast quantities of *training data*. Dataset creators and curators make extensive decisions about how much and which data to include in a training dataset. These choices directly and significantly shape a model's outputs (a.k.a. *generations*), including the model's capacity to learn concepts and produce novel content.
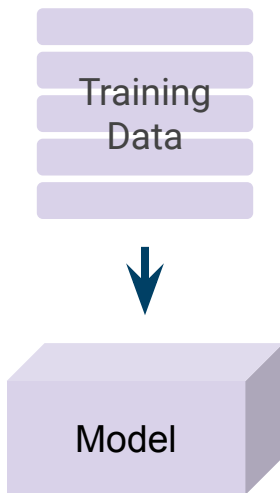
https://genlaw.github.io/explainers/

GenLaw '23
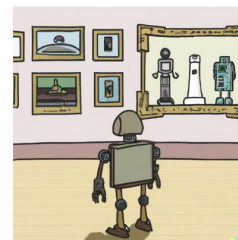
Dataset creation

Model training

Model serving

Model Training

Training Data

Model

Model **training**
involves a lot of
choices

Dataset creation

Model training

Model serving

Dataset creation

Model training

Model serving
Prompting
Generation
Fine-tuning
Human Feedback

Prompting & Generation

Prompt

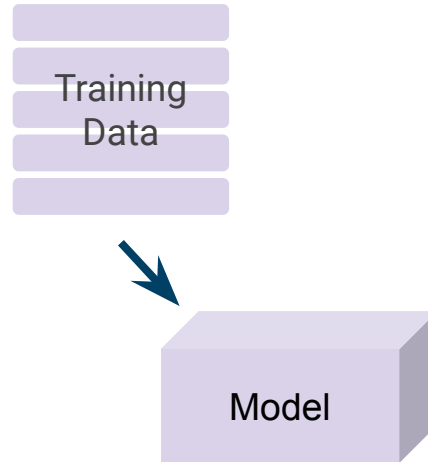Two little robots sitting on the beach reading law books

Model

Generation

Fine-tuning

Training
Data

Model

Fine-tuning

Training Data

Fine-tuning Data

Model

Human Feedback

Prompt

Two little robots sitting on the beach reading law books

Training Data

Fine-tuning Data

Model

Generation
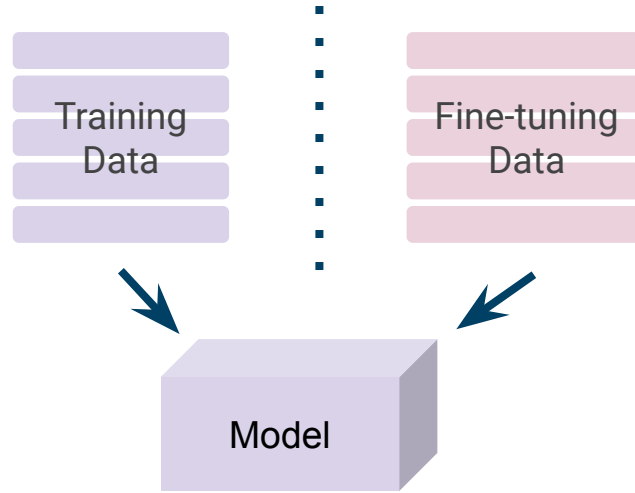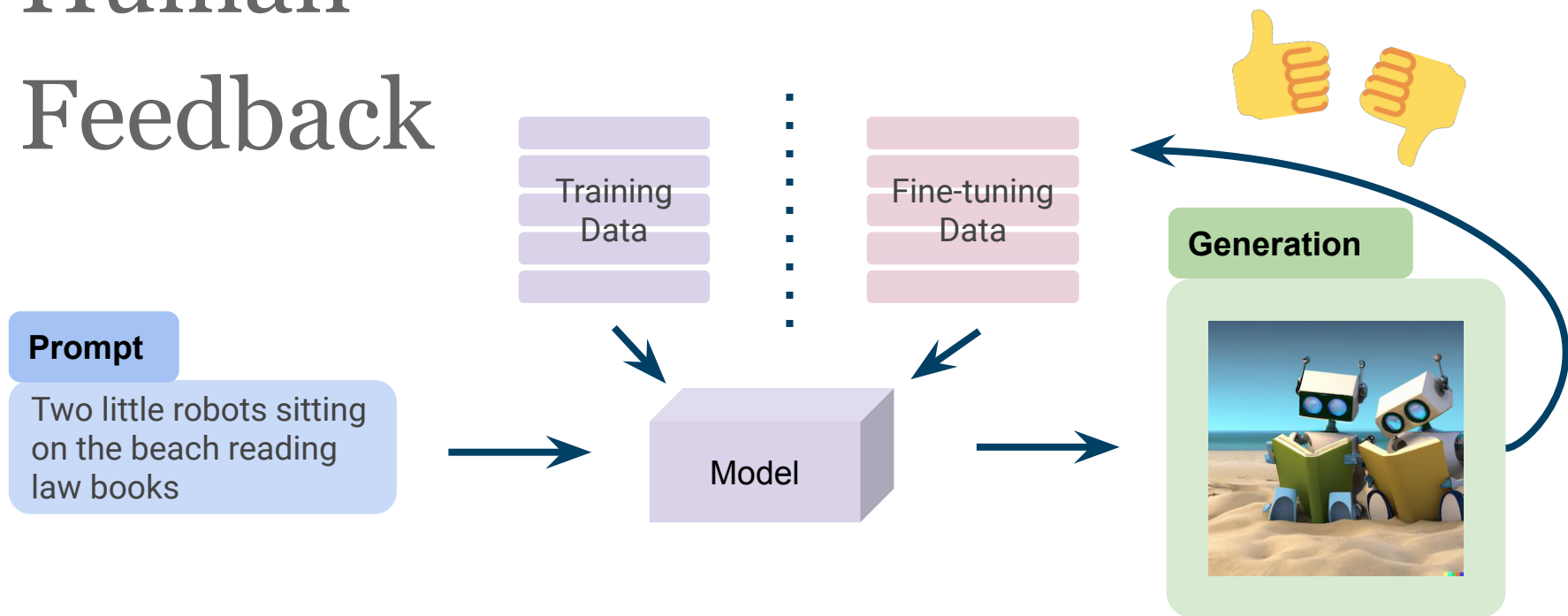
Model **serving** involves a lot of *choices*

Choices

Choices

→

Research Questions

(1) GETTY IMAGES (US), INC.

(a company incorporated under the laws of the State of New York)

(2) GETTY IMAGES INTERNATIONAL

(a company incorporated under the laws of

(3) GETTY IMAGES (UK) LIMITED

(4) GETTY IMAGES DEVCO UK LIMIT

(5) ISTOCKPHOTO LP

J.L., C.B., K.S., P.M., N.G., R.F., J.D. and G.R.

individually, and on behalf of all others situated,

PAUL TREMBLAY, an individual and MONA AWAD, an individual,

Individual and Representative Plainti

SARAH ANDERSEN, an individual; KELLY MCKERNAN, an individual; KARLA ORTIZ, an individual,

Individual and Representative Plaintiffs,

RAH SILVERMAN, an individual; RISTOPHER GOLDEN, an individual; CHARD KADREY, an individual;

Individual and Representative Plaintiffs,

laware nonprofit corporation; OPENA

TIFFS P.M., K.S., B.B., S.J., P., S.A., L.M., D.C., C.L., C R., individually, and on beha ly situated,

v.

STABILITY AI LTD., a UK corporation; STABILITY AI, INC., a Delaware corporation; MIDJOURNEY, INC., a Delaware corporation; DEVIANTART, INC., a Delaware corporation,

Defendants.

ENAI, INC., a Delaware nonprofit corporation; OPENAI, L.P., a aware limited partnership; OPENAI OPCO, L.L.C., a Delaware ited liability corporation; OPENAI GP, L.L.C., a Delaware ited liability company; OPENAI STARTUP FUND GP I, L.L.C., limited liability company; OPENAI STARTUP FUND I, aware limited partnership; and OPENAI STARTUP FUND MENT, LLC, a Delaware limited liability company,

Defendants.

Plaintiffs,

VS.

AI LP, OPENAI INCORPO AI GP, LLC, OPENAI STA OPENAI STARTUP FUND OPENAI STARTUP FUND MAN LLC, MICROSOFT CORPORATI 1 through 20, inclusive,

Defendants.

v.

META PLATFORMS, INC., a Delaware

;

oration; OPENAI, INC., a Delawar profit corporation; OPENAI, L.P., a aware limited partnership; OPENAI C., a Delaware limited liability com NAI STARTUP FUND GP L.L

Choices → Research Questions

IP + Privacy

Organizers

Katherine Lee

Ph.D. Candidate

Cornell University Department
of Computer Science

A. Feder Cooper

Ph.D. Candidate

Cornell University Department
of Computer Science

**Fatemehsadat
Mireshghallah**

Post-Doctoral Researcher

University of Washington

Madiha Z. Choksi

Ph.D. Student

Cornell University Department
of Information Science

James Grimmelmann

*Professor of Digital and
Information Law*

Cornell Law School and
Cornell Tech

David Mimno

Associate Professor

Cornell University Department
of Information Science

Deep Ganguli

Research Scientist

Anthropic

Ludwig Schubert

GenLaw '23

Organizers

Katherine Lee

Ph.D. Candidate

Cornell University Department
of Computer Science

A. Feder Cooper

Ph.D. Candidate

Cornell University Department
of Computer Science

**Fatemehsadat
Mireshghallah**

Post-Doctoral Researcher

University of Washington

Madiha Z. Choksi

Ph.D. Student

Cornell University Department
of Information Science

James Grimmelmann

*Professor of Digital and
Information Law*

Cornell Law School and
Cornell Tech

David Mimno

Associate Professor

Cornell University Department
of Information Science

Deep Ganguli

Research Scientist

Anthropic

Ludwig Schubert

GenLaw '23

Time	Speaker
9:15 - 9:45 AM	Pamela Samuelson, "Some Nonobvious Observation About Copyright's Scope For Generative AI Developers"
9:45 - 10:05 AM	Mark Lemley, "Is Training AI Copyright Infringement?"
10:05 - 10:40 AM	**Coffee**
10:40 - 11:00 AM	Miles Brundage, "Where and when does the law fit into AI development and deployment?"
11:00 - 12:00 PM	*Panel on Intellectual Property*: Pamela Samuelson, Mark Lemley, Luis Villa, Katherine Lee (Moderated by Jack Balkin and A. Feder Cooper)
12:00 - 1:30 PM	**Lunch**
1:30 - 1:45 PM	Jack Balkin, "Artificial Intelligence and the First Amendment"
1:45 - 2:15 PM	**Spotlights**
2:15 - 3:00 PM	**Posters**
3:00 - 3:30 PM	**Coffee**
3:30 - 3:45 PM	Nicholas Carlini, "A Brief Introduction to Machine Learning & Memorization"
3:45 - 4:00 PM	Gautam Kamath, "What does Differential Privacy have to do with Copyright?"
4:00 - 5:00 PM	*Panel on Privacy*: Kristen Vaccaro, Nicholas Carlini, Miles Brundage, Gautam Kamath, and Jack Balkin (Moderated by Katherine Lee and Deep Ganguli)

Generative AI + Law

GenLaw '23

Speakers:

Pamela Samuelson

Distinguished Professor of Law and Information

University of California, Berkeley

Mark Lemley

Professor of Law

Stanford Law School

Luis Villa

Co-founder and General Counsel

Tidelift

Katherine Lee

Ph.D. Candidate

Cornell University Department of Computer Science

Moderators:

Jack M. Balkin

Professor

Yale Law School

A. Feder Cooper

Ph.D. Candidate

Cornell University Department of Computer Science

GenLaw '23

Speakers:

Kristen Vaccaro

Assistant Professor

University of California, San Diego

Nicholas Carlini

Research Scientist

Google DeepMind

Gautam Kamath

Assistant Professor

University of Waterloo

Jack M. Balkin

Professor

Yale Law School

Miles Brundage

Head of Policy Research

OpenAI

Moderators:

Katherine Lee

Ph.D. Candidate

Cornell University Department of Computer Science

Deep Ganguli

Research Scientist

Anthropic

GenLaw '23

Takeaways

Interpretation
Evaluation
Alignment

Takeaways

Interpretation
Evaluation
Alignment

IP + Privacy