CALIFORNIA
LAWYERS
ASSOCIATION

California Lawyers Association

presents

Can ChatGPT Do That? How to Embed Human Values into AI: A Lawyer's Guide to Privacy, Ethics and Software Development

1.25 Hours MCLE; 1.25 Legal Ethics

Saturday, September 23, 2023

11:30 AM -12:45 PM

Speakers:

Jeeyun (Sophia) Baik

Serge Egelman

Jeewon Kim Serrato

Conference Reference Materials

Points of view or opinions expressed in these pages are those of the speaker(s) and/or author(s). They have not been adopted or endorsed by the California Lawyers Association and do not constitute the official position or policy of the California Lawyers Association. Nothing contained herein is intended to address any specific legal inquiry, nor is it a substitute for independent legal research to original sources or obtaining separate legal advice regarding specific legal situations.

The California Lawyers Association is an approved State Bar of California MCLE provider.

Can ChatGPT Do That? How to Embed Human Values into AI: A Lawyer’s Guide to Privacy, Ethics and Software Development

Jeeyun (Sophia) Baik, University of San Diego
Serge Egelman, University of California at Berkeley
Jeewon Kim Serrato, BakerHostetler

SAN DIEGO / SEPTEMBER 21- 23

2023 ANNUAL MEETING

BREAKING BARRIERS

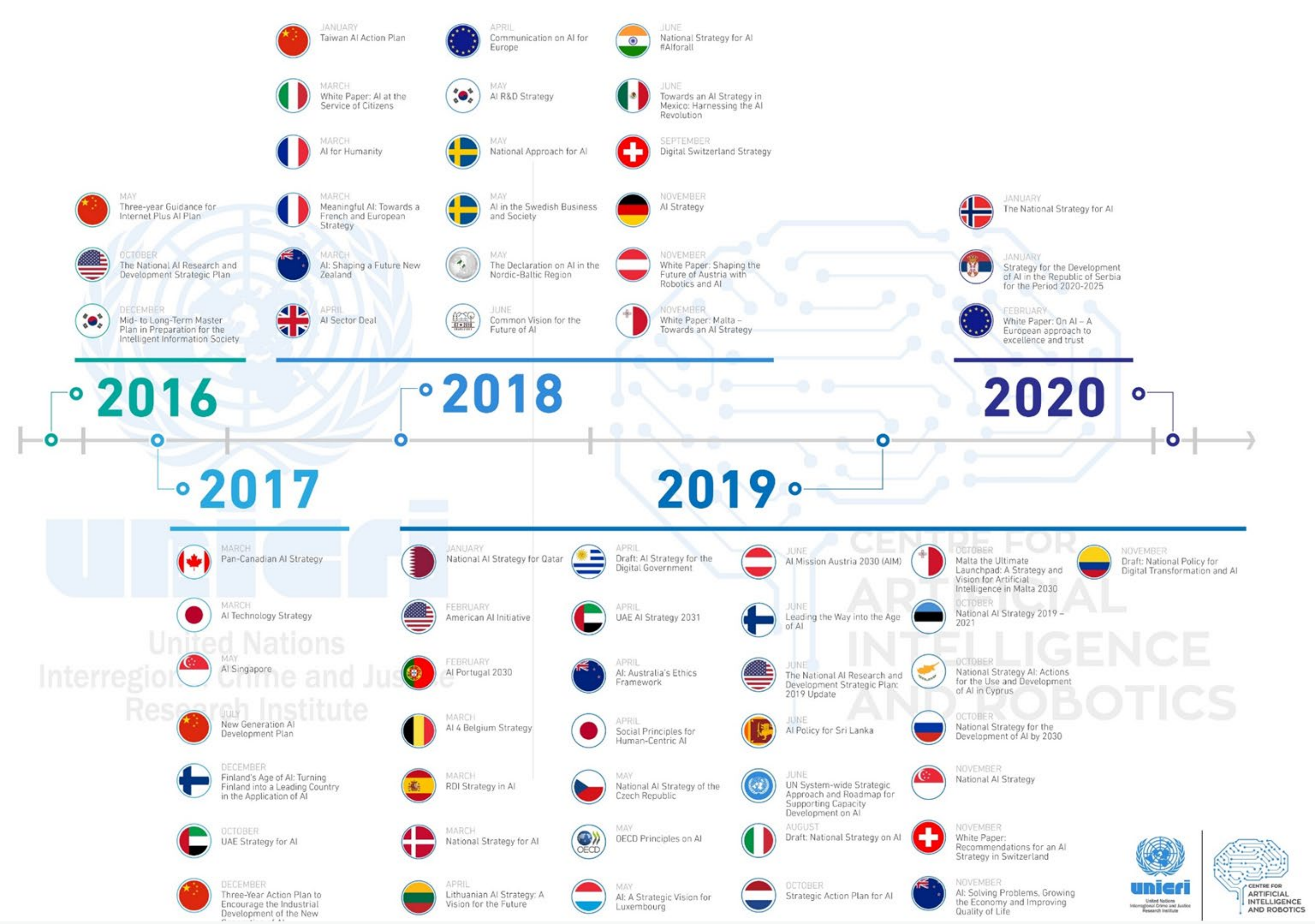
CALIFORNIA LAWYERS ASSOCIATION

#CLAAnnual

Issues

1. Emerging data privacy laws impacting the use of AI products and services in the US and globally
2. Regulations of the use of automated decision-making and digital surveillance
3. How businesses can embed human values and ethics into AI-powered software

National AI Strategies, Action Plans, and Proposals



2016

MAY
Three-year Guidance for Internet Plus AI Plan

OCTOBER
The National AI Research and Development Strategic Plan

DECEMBER
Mid- to Long-Term Master Plan in Preparation for the Intelligent Information Society

2017

MARCH
Pan-Canadian AI Strategy

MARCH
AI Technology Strategy

MAY
AI Singapore

JULY
New Generation AI Development Plan

DECEMBER
Finland's Age of AI: Turning Finland into a Leading Country in the Application of AI

OCTOBER
UAE Strategy for AI

DECEMBER
Three-Year Action Plan to Encourage the Industrial Development of the New

2018

JANUARY
Taiwan AI Action Plan

MARCH
White Paper: AI at the Service of Citizens

MARCH
AI for Humanity

MARCH
Meaningful AI: Towards a French and European Strategy

MARCH
AI: Shaping a Future New Zealand

APRIL
AI Sector Deal

APRIL
Communication on AI for Europe

MAY
AI R&D Strategy

MAY
National Approach for AI

MAY
AI in the Swedish Business and Society

MAY
The Declaration on AI in the Nordic-Baltic Region

JUNE
Common Vision for the Future of AI

JUNE
National Strategy for AI #AIforall

JUNE
Towards an AI Strategy in Mexico: Harnessing the AI Revolution

SEPTEMBER
Digital Switzerland Strategy

NOVEMBER
AI Strategy

NOVEMBER
White Paper: Shaping the Future of Austria with Robotics and AI

NOVEMBER
White Paper: Malta – Towards an AI Strategy

2019

JANUARY
National AI Strategy for Qatar

FEBRUARY
American AI Initiative

FEBRUARY
AI Portugal 2030

MARCH
AI 4 Belgium Strategy

MARCH
RDI Strategy in AI

MARCH
National Strategy for AI

APRIL
Lithuanian AI Strategy: A Vision for the Future

APRIL
Draft: AI Strategy for the Digital Government

APRIL
UAE AI Strategy 2031

APRIL
AI: Australia's Ethics Framework

APRIL
Social Principles for Human-Centric AI

MAY
National AI Strategy of the Czech Republic

MAY
OECD Principles on AI

MAY
AI: A Strategic Vision for Luxembourg

JUNE
AI Mission Austria 2030 (AIM)

JUNE
Leading the Way into the Age of AI

JUNE
The National AI Research and Development Strategic Plan: 2019 Update

JUNE
AI Policy for Sri Lanka

JUNE
UN System-wide Strategic Approach and Roadmap for Supporting Capacity Development on AI

AUGUST
Draft: National Strategy on AI

OCTOBER
Strategic Action Plan for AI

OCTOBER
Malta the Ultimate Launchpad: A Strategy and Vision for Artificial Intelligence in Malta 2030

OCTOBER
National AI Strategy 2019 – 2021

OCTOBER
National Strategy AI: Actions for the Use and Development of AI in Cyprus

OCTOBER
National Strategy for the Development of AI by 2030

NOVEMBER
National AI Strategy

NOVEMBER
White Paper: Recommendations for an AI Strategy in Switzerland

NOVEMBER
AI: Solving Problems, Growing the Economy and Improving Quality of Life

NOVEMBER
Draft: National Policy for Digital Transformation and AI

2020

JANUARY
The National Strategy for AI

JANUARY
Strategy for the Development of AI in the Republic of Serbia for the Period 2020-2025

FEBRUARY
White Paper: On AI – A European approach to excellence and trust

unicri
United Nations Interregional Crime and Justice Research Institute
CENTRE FOR ARTIFICIAL INTELLIGENCE AND ROBOTICS

AI Framework Considerations

- **Beneficence**: Promoting well-being, preserving dignity, and sustaining the planet
- **Non-maleficence**: Privacy, security and “capability caution”
- **Autonomy**: Power to decide (to decide)
- **Justice**: Promoting prosperity, preserving solidarity, avoiding unfairness
- **Explicability**: Enabling the other principles through intelligibility and accountability

US Approach to AI

Federal

- Section 5 FTC Act
- Fair Credit Reporting Act
- Equal Credit Opportunity Act
- FTC April 2020 Guidance "Using Artificial Intelligence and Algorithms"
- FTC January 2016 Report "Big Data: A Tool for Inclusion or Exclusion?"
- FTC September 2014 "Big Data" Workshop on data modeling, data mining, and analytics
- U.S. Dept. of Commerce, National AI Advisory Committee
- NIST Special Publication Standard for Identifying and Managing Bias in Artificial Intelligence
- White House Blueprint for an AI Bill of Rights

State and Local

- California Consumer Privacy Act (CCPA), amended by California Privacy Rights Act (CPRA)
- Colorado Privacy Act (CPA)
- Connecticut Privacy Act (CTPA)
- Virginia Consumer Data Protection Act (VCDPA)
- New York City Local Law Int. No. 1894-A Regulating the Use of Artificial Intelligence in Employment Decisions

US Right to Object to Automated Decision-making

Regulations Pending

- Automated decisionmaking technology
- Profiling
- Algorithmic Discrimination
- Access and/or Opt-Out Rights in the Context of Automated Decisionmaking
- Legal or Similarly Significant Effects Concerning a Consumer
- Human Involved Automated Processing, Human Reviewed Automated Processing, and Solely Automated Processing

Existing Requirements

- Notice at Collection
- Right to Access
- Consent
- Purpose Limitation
- Contracts

Defining Automated Decision-making

CCPA: access and opt-out rights regarding use of “any system, software, or process–including one derived from machine-learning, statistics, or other data processing or artificial intelligence techniques–that processes personal information and uses computation as whole or part of a system to make or execute a decision or facilitate human decision making” (as recommended by CPPA subcommittee)

EU’s GDPR: “The data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her.”

Profiling as a Key Part of Automated Decision-making

CCPA: “any form of automated processing of personal information…to evaluate certain personal aspects relating to a natural person and in particular to analyze or predict aspects concerning that natural person’s performance at work, economic situation, health, personal preferences, interests, reliability, behavior, location, or movements.”

EU’s GDPR: “The data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her.”

Scoping the Harm: (1) Human Involvement

CPA (Colorado Privacy Act):

- "Businesses should honor the opt-out request of solely automated and human reviewed automated processing."
- "Businesses should follow certain procedures if a request to opt-out of human involved automated processing is denied."

EU's GDPR: "The data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her."

Scoping the Harm: (2) Types of Risks

EU's GDPR:

"A decision produces legal effects when your legal rights are impacted (such as your right to vote). In addition, processing can significantly affect you if it influences your circumstances, behaviour or choices. For example automatic processing may lead to the refusal of your online credit application." (European Commission)

EU's AI Act:

- Unacceptable Risk ⇒ banned (e.g., social scoring, real-time facial recognition)
- High Risk ⇒ assessed before a market launch and throughout its lifecycle (e.g., AI systems used for employment, border control, law enforcement)
- Generative AI ⇒ transparency requirements (e.g., ChatGPT)
- Limited Risk ⇒ minimal transparency requirements (e.g., image-generating app)

(European Parliament)

Ensuring Users' Awareness of AI in Use

EU's AI Act

a. Generative AI, like ChatGPT ⇒ "Disclosing that the content was generated by AI"
b. Limited Risk AI ⇒ "Users should be made aware when they are interacting with AI"

(European Parliament)

Risk Assessment

CPA (Colorado Privacy Act):

- "Companies should conduct data protection assessments for processing activities that present a heightened risk of harm."
- "Companies should update its assessment at minimum whenever the level of risk presented by an existing processing activity is materially modified…"

CCPA (California Consumer Privacy Act):

- CPRA proposed to require "businesses whose processing of consumers' personal information presents significant risk to consumers' privacy or security, to…perform a cybersecurity audit on an annual basis…[and to] submit to the California Privacy Protection Agency on a regular basis a risk assessment…"
- CPPA is developing the standards through a rulemaking process.

Training Data

Provenance:

- Where is training data coming from?
- Is it allowed to be used for this purpose?
- Are there applicable laws?
- What were data subjects told?
- Did data subjects consent?

Use of 3rd-party services presents additional issues:

- Is sensitive data being used to trained models used by others?
- Do you disclose this to your users?

Accountability & Fairness

Accuracy:

- How accurate is the data?
- How might inaccurate data impact results?
- How might inaccurate data lead to biased results?
- What risks does this create for the organization?

Organizational Pitfalls

Most software privacy issues are due to mistakes that led to data being collected in unanticipated ways.

Regarding AI, it’s critical that legal teams regularly communicate with engineering teams to:

- stay up-to-date on how the organization may be using AI to make decisions that impact others (e.g., employees, customers, the public, etc.)
- understand when personal data is being used to train AI models
- understand when personal or other sensitive data is being used by service providers to train their own AI models

Case Study 1: GDPR’s Approach to Targeted Advertising

“[T]he personalised advertising by which the online social network Facebook finances its activity, cannot justify, as a legitimate interest pursued by Meta Platforms Ireland, the processing of the data at issue, in the absence of the data subject’s consent”

(July 4, 2023 - Court of Justice of the European Union)

Pitfalls of Loyalty Programs

Case Study 2: CPPA’s Review of Connected Vehicles

Modern vehicles are effectively connected computers on wheels. They’re able to collect a wealth of information via built-in apps, sensors, and cameras, which can monitor people both inside and near the vehicle.

- Ashkan Soltani, CPPA’s Executive Director

Case Study 3: FTC AI Enforcement Cases

- **FTC** can enforce Section 5 of the FTC Act, FCRA, and ECOA against users and developers of unfair or biased algorithms.
- **FTC** ordered a company to delete models and algorithms developed using users’ uploaded photos and videos.
- **FTC** ordered a company to destroy any models or algorithms developed with the use of improperly collected children’s personal information.

Takeaways

1. Need to think about whether and how to address the level of human involvement in any AI systems
2. Should be proactive in ensuring users' comprehensive awareness of an AI system in use, beyond a one-time notice-and-choice
3. Must pay attention to the regulatory development of risk assessment requirements and how an AI system's risk level can change and evolve.