SOLASAI

Defining and Measuring Fairness

Northeastern University: Big Data and Intelligent Analytics

Nicholas Schmidt

November 11, 2022

Outline

- Introduction
- Why Should We Care About Model Fairness?
- Frameworks for Understanding Discrimination
- Making Models Fairer
- Code Demonstration: Testing Models for Evidence of Discrimination

A little bit About Me

- **Nicholas Schmidt**
    - 20+ years of experience applying concepts from statistics and economics to questions of law and regulatory compliance.

- **CEO, SolasAI**
    - SolasAI software ***measures*** and ***mitigates*** discrimination risk.
    - Prominent U.S. lenders, insurers, and health insurance companies are using SolasAI to assess and mitigate discrimination risk.

- **AI Practice Leader, BLDS, LLC**
    - We are the fair lending analytics advisors to lenders that represent over 70% of credit cards issued in the United States.
    - We are regularly engaged by regulators and courts to provide guidance on discrimination risk in algorithms.

Can ai discriminate?

1%

Gender was misidentified in **up to 1 percent of lighter-skinned males** in a set of 385 photos.

Gender was misidentified in **up to 12 percent of darker-skinned males** in a set of 318 photos.

12%

7%

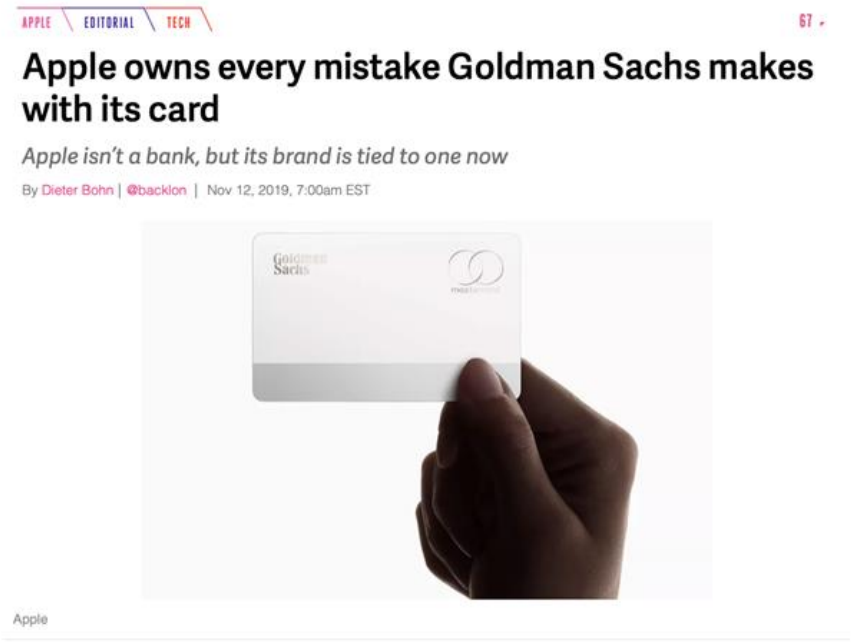
Gender was misidentified in **up to 7 percent of lighter-skinned females** in a set of 296 photos.

Gender was misidentified in **35 percent of darker-skinned females** in a set of 271 photos.

35%

Lohr, Steve. "Facial recognition is accurate, if you're a white guy." *New York Times*, 9 February 2018.

The effect of two people on twitter



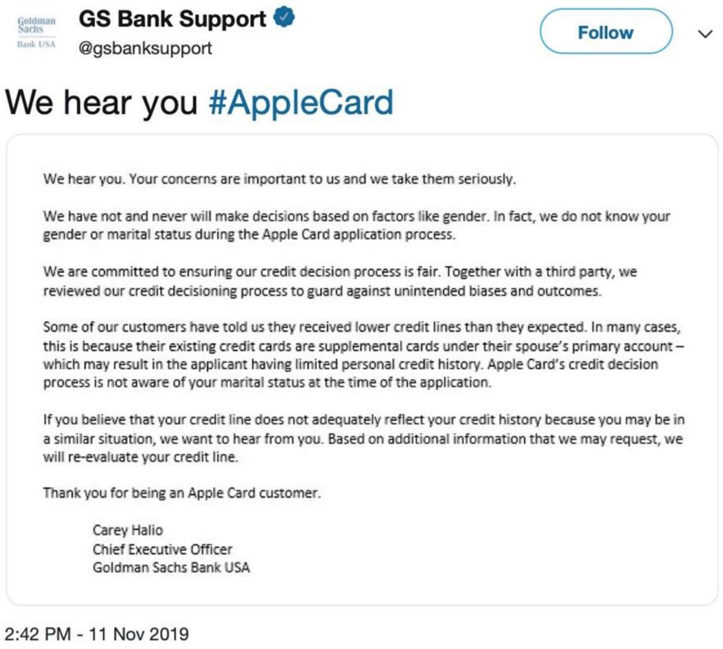
APPLE EDITORIAL TECH 67

Apple owns every mistake Goldman Sachs makes with its card

Apple isn't a bank, but its brand is tied to one now

By Dieter Bohn | @backlon | Nov 12, 2019, 7:00am EST

Apple



GS Bank Support
@gsbanksupport

We hear you #AppleCard

We hear you. Your concerns are important to us and we take them seriously.

We have not and never will make decisions based on factors like gender. In fact, we do not know your gender or marital status during the Apple Card application process.

We are committed to ensuring our credit decision process is fair. Together with a third party, we reviewed our credit decisioning process to guard against unintended biases and outcomes.

Some of our customers have told us they received lower credit lines than they expected. In many cases, this is because their existing credit cards are supplemental cards under their spouse's primary account – which may result in the applicant having limited personal credit history. Apple Card's credit decision process is not aware of your marital status at the time of the application.

If you believe that your credit line does not adequately reflect your credit history because you may be in a similar situation, we want to hear from you. Based on additional information that we may request, we will re-evaluate your credit line.

Thank you for being an Apple Card customer.

Carey Halio
Chief Executive Officer
Goldman Sachs Bank USA

2:42 PM - 11 Nov 2019

Frameworks for Understanding Types of Discrimination and Bias

Conceptual Framework

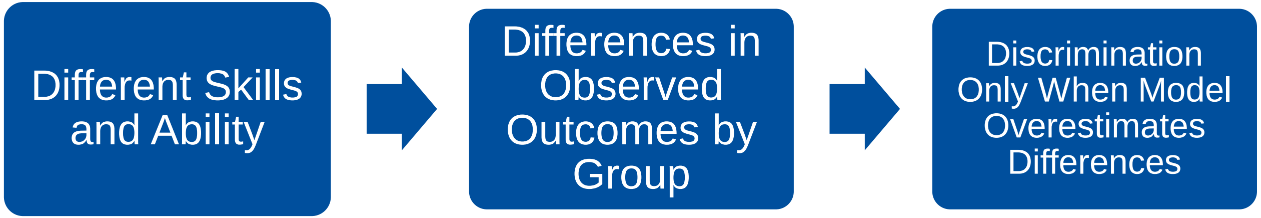
- **Outlook**
  - What You See is What You Get
  - We Are All Equal
- **Measurement (Affects)**
  - Groups
  - Individuals

Legal Framework

- **Disparate Treatment**
  - Explicit (even unintentional) consideration of characteristics
- **Disparate Impact**
  - Factors with a valid but discriminatory effect
- **Proxy Discrimination**
  - Factor that nearly identifies group membership

SOLASAI

Concepts of Fairness: *What You See is What You Get*

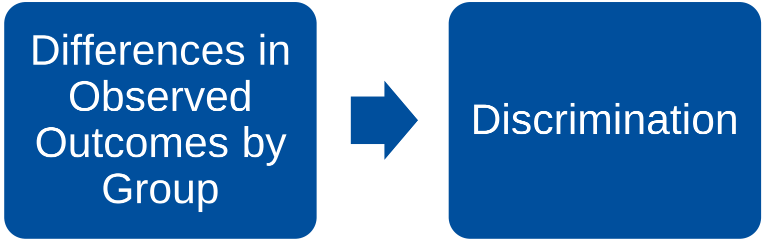


Different Skills and Ability

Differences in Observed Outcomes by Group

Discrimination Only When Model Overestimates Differences

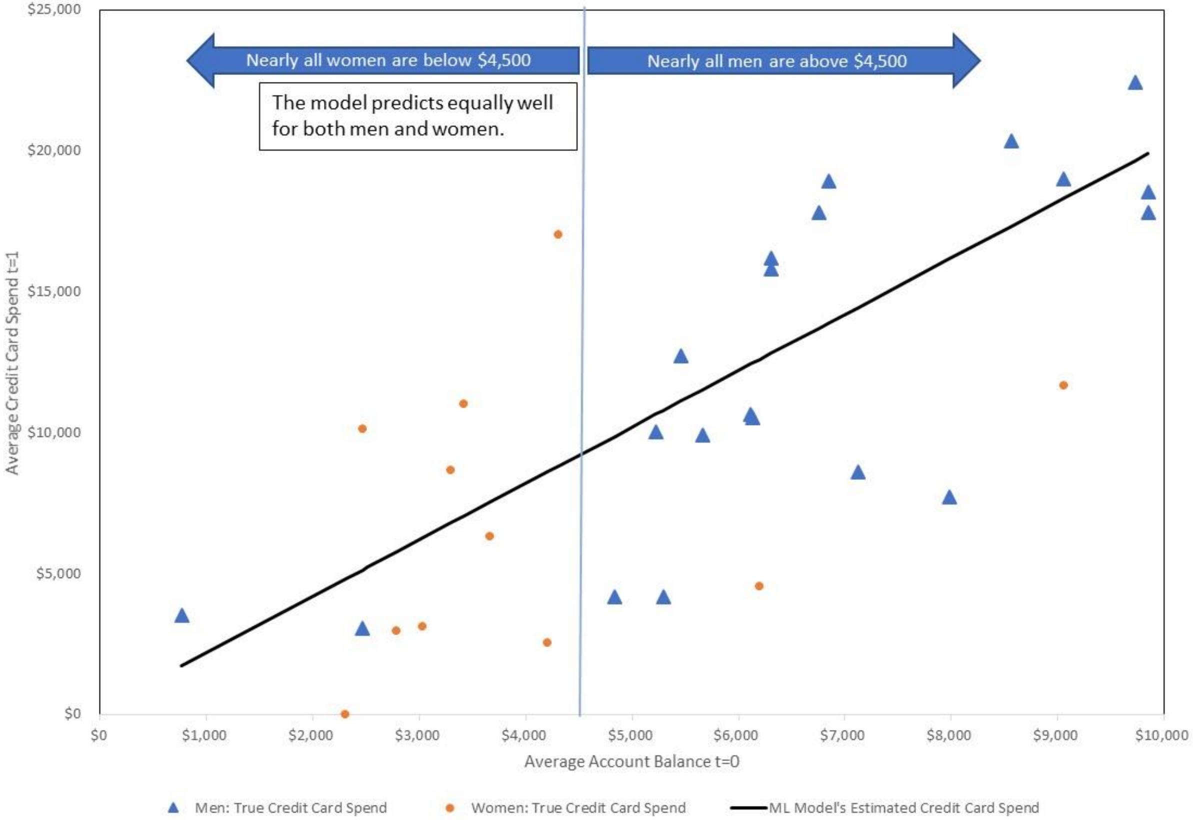
Concepts of Fairness: *We are All Equal*



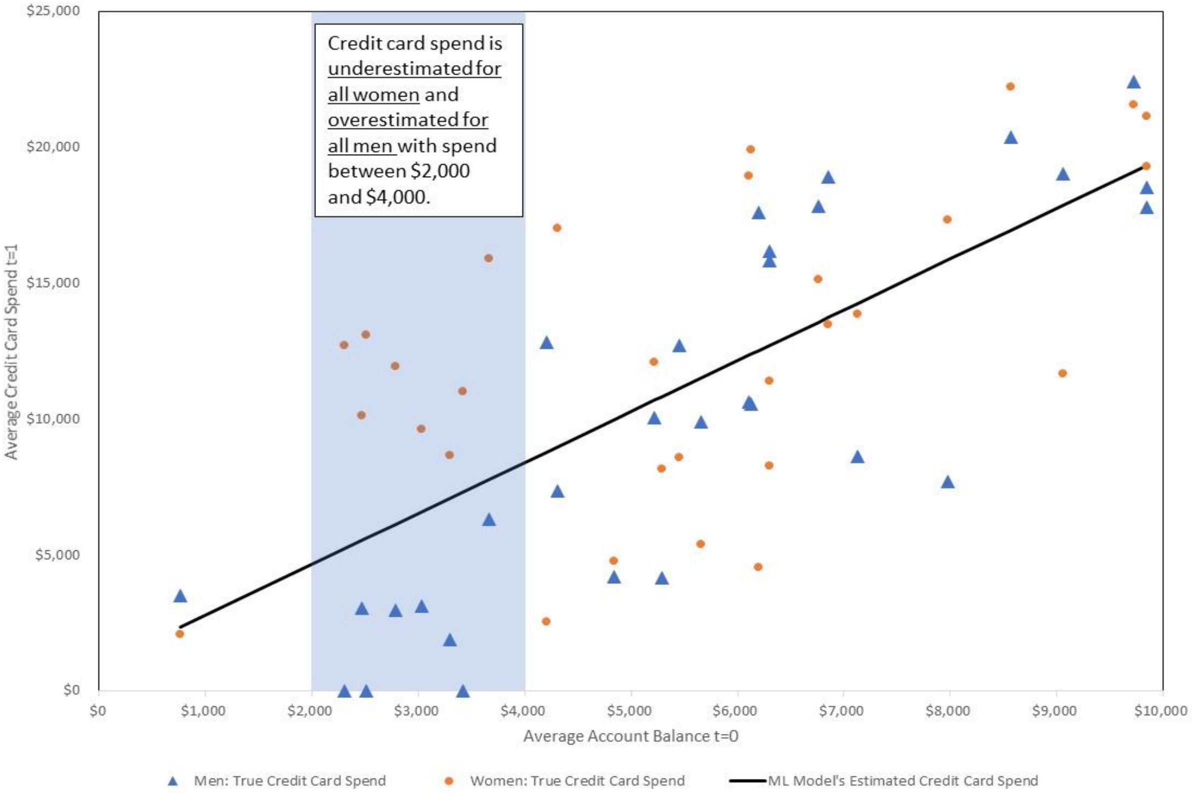
Differences in Observed Outcomes by Group

Discrimination

Disparate Impact

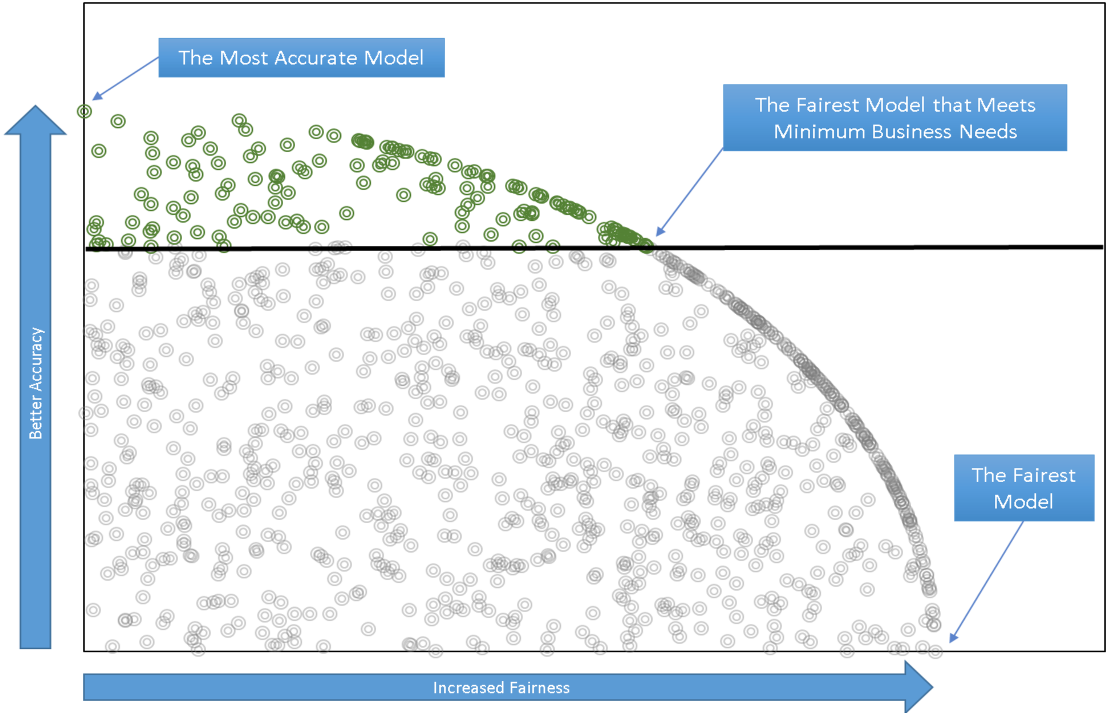


Differential Validity



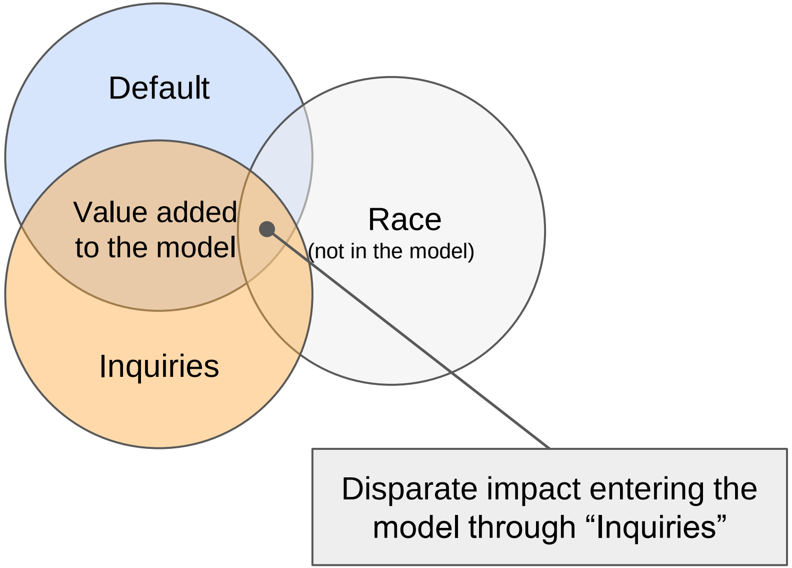
Credit card spend is underestimated for all women and overestimated for all men with spend between $2,000 and $4,000.

Using AI to Fix AI: the Pareto Frontier



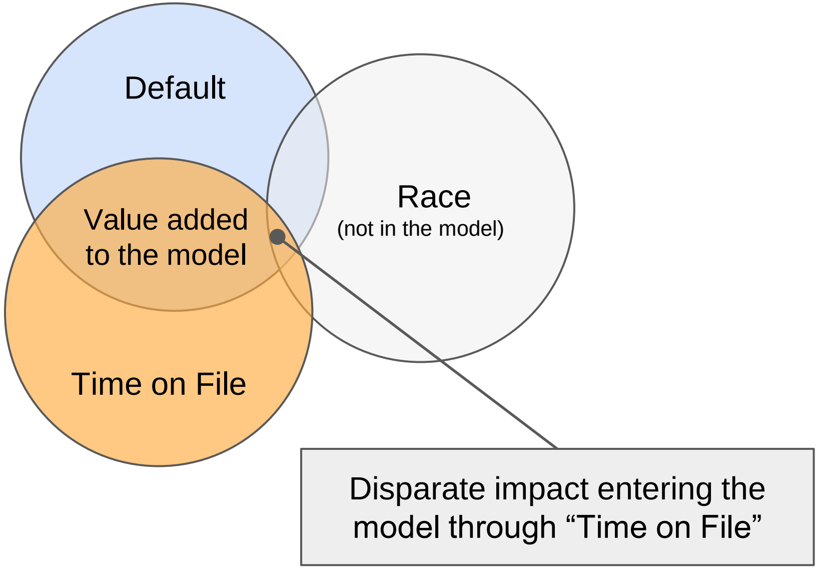
Making Fairer Models – Feature Selection

Model 1 - Includes Inquiries



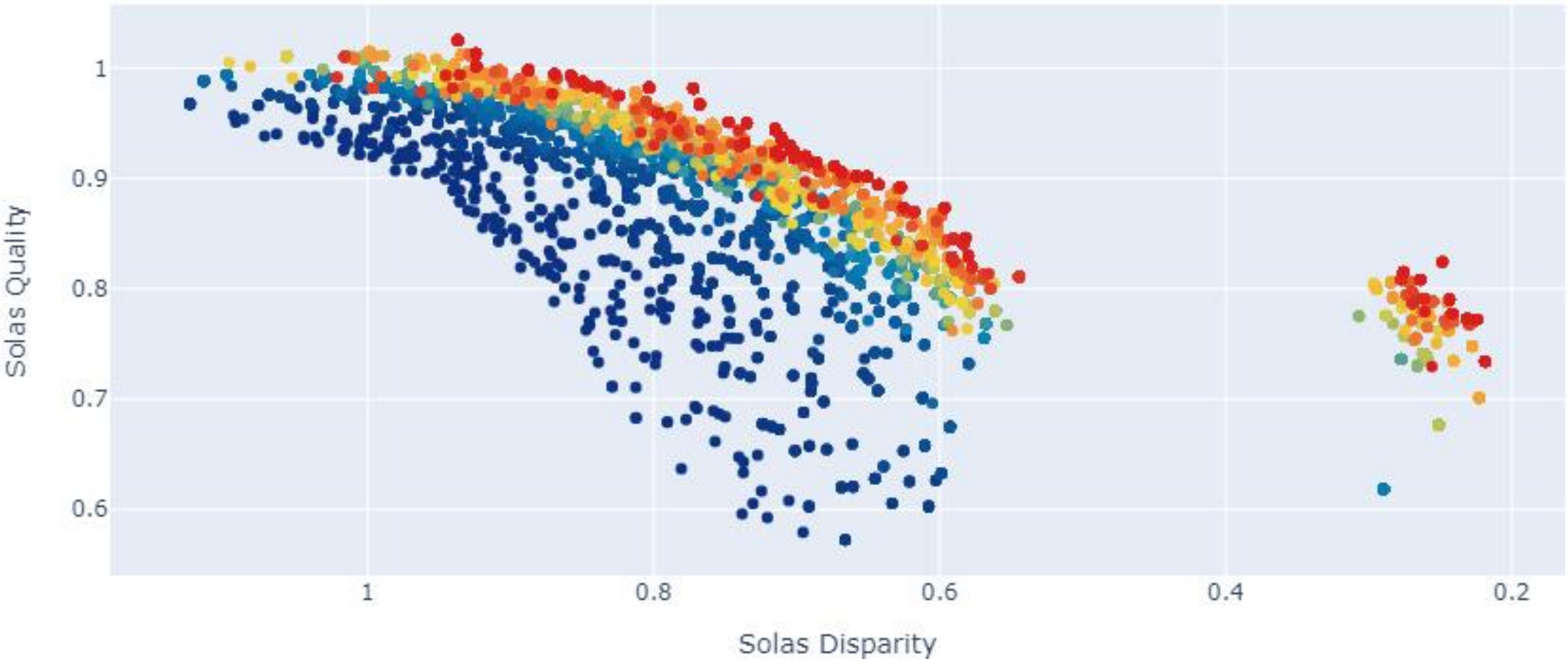
Disparate impact entering the model through “Inquiries”

Model 2 - Includes Time on File



Disparate impact entering the model through “Time on File”

Using AI to Fix AI: A Real-world Example



Measuring Discrimination / Code Demonstration

$$Adverse\ Impact\ Ratio\ (AIR) = \frac{\%\ Protected\ Group\ Selected}{\%\ Reference\ Group\ Selected}$$

$$Standardized\ Mean\ Difference\ (SMD) = 100 * \left(\frac{\hat{Y}_{protected\ group} - \hat{Y}_{reference\ group}}{\sigma_{\hat{Y}}} \right)$$

$$Residual\ SMD\ (rSMD) = 100 * \left(\frac{\varepsilon_{protected\ group} - \varepsilon_{reference\ group}}{\sigma_{\varepsilon}} \right)$$

- Available on GitHub: https://github.com/nickpschmidt/public_talks

SOLASAI

Thank You

Nicholas Schmidt
nicholas.schmidt@solas.ai
https://www.linkedin.com/in/nickpschmidt