The Challenge of Trusted AI

Marco Hernansanz

EVP & CEO Southern Europe, Middle East and Africa

Salesforce

The Headlines Tell A Story

Generative AI will change the world – Machine vs. Human Edition

INSIDER

ChatGPT is a mind-blowing 'game changer' that feels like magic, says Coursera CEO

Using Solar-Powered AI Sensors to Detect Wildfires Earlier

CNET

The New York Times

A Stroke Stole Her Ability to Speak at 30. A.I. Is Helping to Restore It Years Later.

But the Headlines Also Tell Another Story

WIRED

AI Has a Hallucination Problem That's Proving Tough to Fix

npr

Google shares drop $100 billion after its new AI chatbot makes a mistake

Vox

AI's Islamophobia problem

GPT-3 is a smart and poetic AI. It also says terrible things about Muslims.

AXIOS

Companies are struggling to keep corporate secrets out of ChatGPT

FAST COMPANY

Why Amazon's 'dead grandma' Alexa is just the beginning for voice cloning

VentureBeat

How ChatGPT can turn anyone into a ransomware and malware threat actor

The Brussels Times

Belgian man dies by suicide following exchanges with chatbot

Organizations will only adopt AI if it's trustworthy

1

Accuracy and Confabulations

2

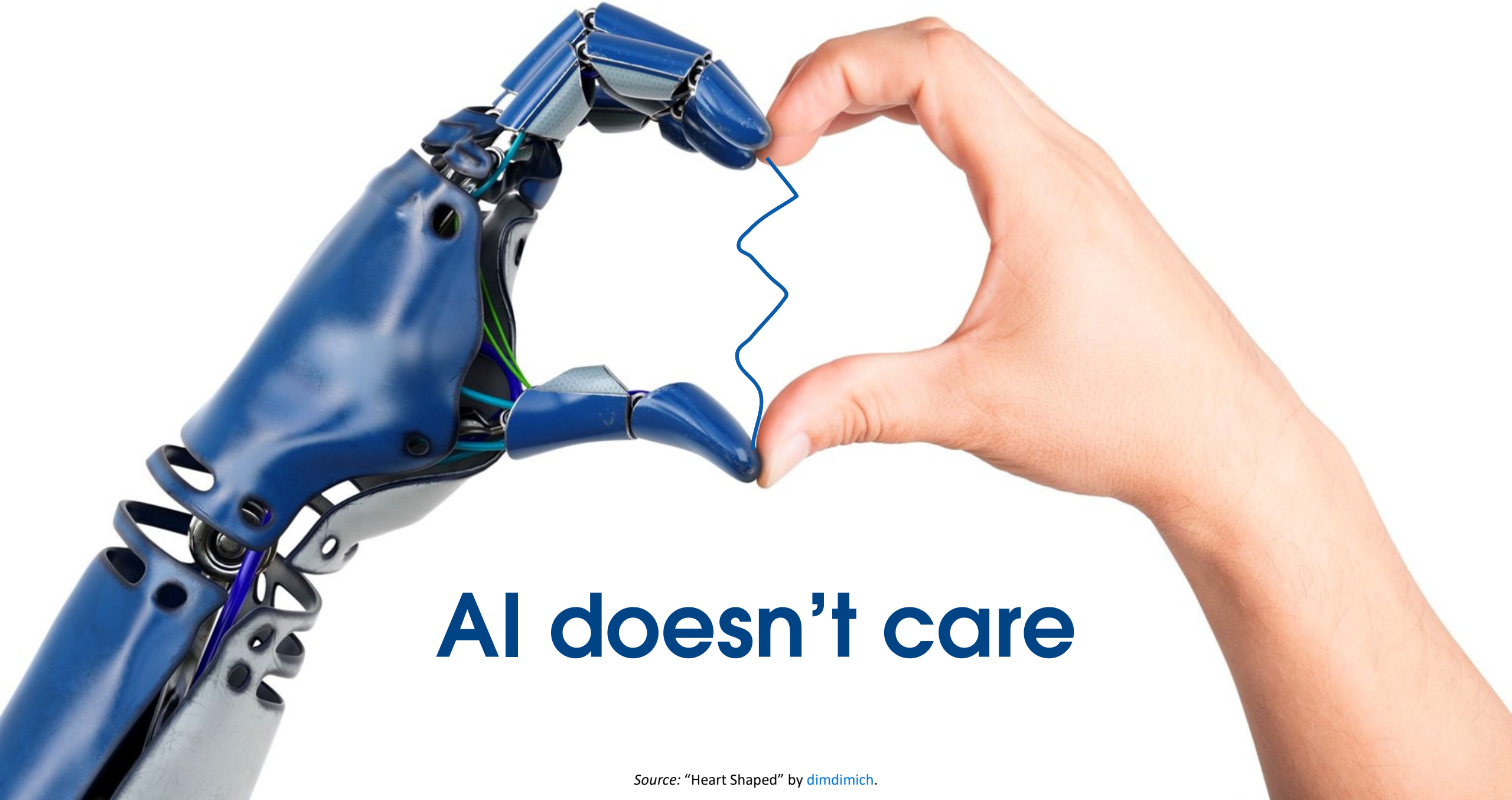
Bias and Fairness

3

Privacy and Security

4

Job Impact or Equity to Access



AI doesn't care

Source: "Heart Shaped" by dimdimich.

That's why we have to focus on a trustworthy AI…

Principles

Policies

Protections

5 Guidelines for Responsible Generative AI

sfdc.co/RGAI-guidelines

Accurate

Be accurate and convey uncertainty when the answer isn't clear.

Enable fact-checking when possible.

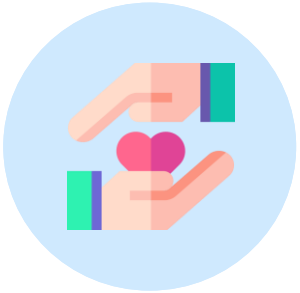
Safe

Mitigate bias, toxicity, and harmful content.

Protect PII and prevent data leakage

Honest

Respect data provenance and make clear that content is AI-produced when autonomously delivered.

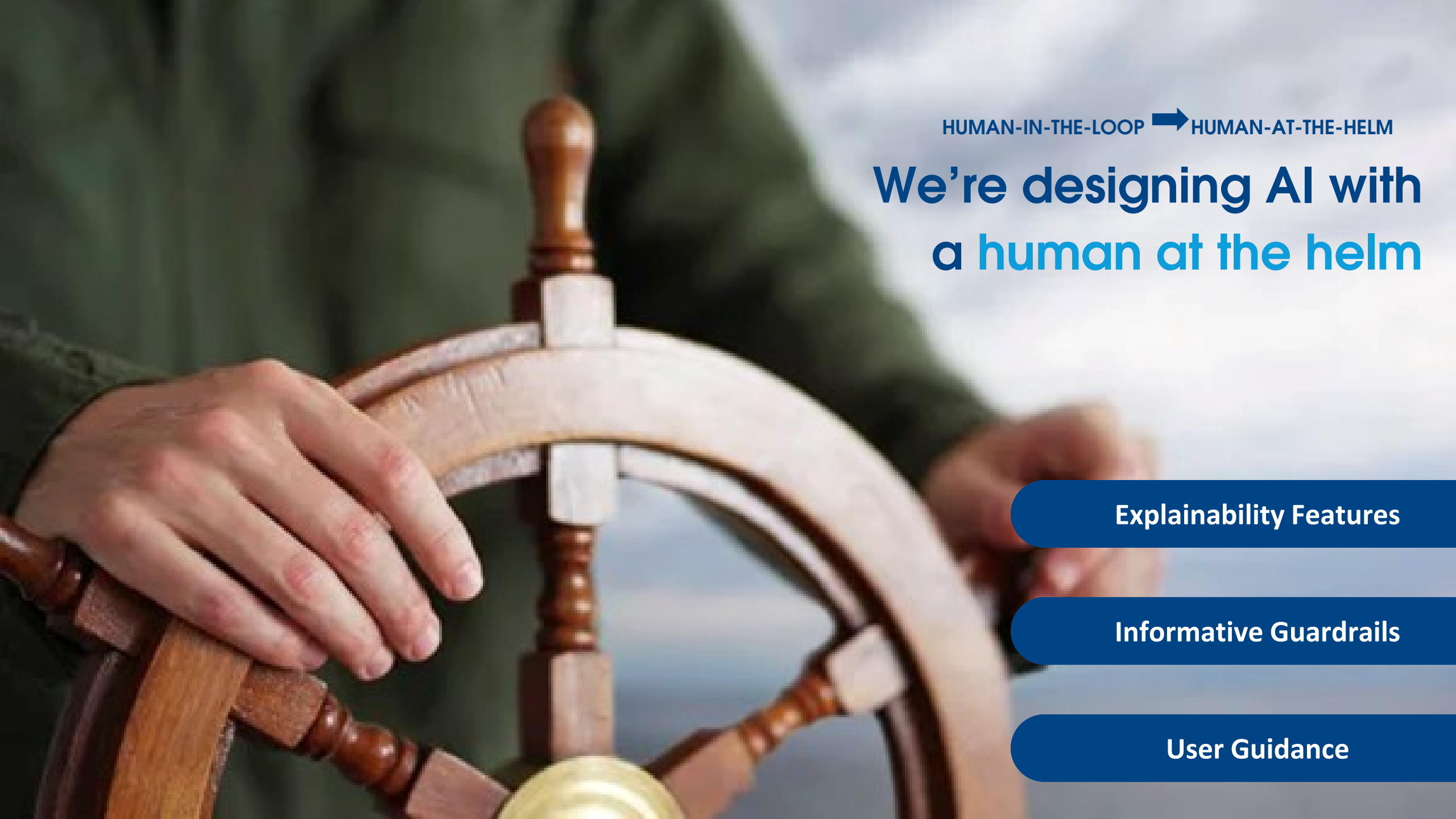


Empowering

Supercharge human capabilities, accessible to all, and engage in responsible labor practices.

Sustainable

Right-size models to reduce carbon and water footprint.



HUMAN-IN-THE-LOOP ➡ HUMAN-AT-THE-HELM

We're designing AI with a human at the helm

- **Explainability Features**
- **Informative Guardrails**
- **User Guidance**

“Trust is a confident relationship with the unknown”

-Rachel Botsman, author & expert on trusted technology