Where: [Q-Branch, Downtown Austin](#)

When: Dec 4, 2024, 6PM

Who: 100-125 Austin AI Alliance Members, select clients / partners of IBM

Volunteers:

- Q branch has elevators guides
- Sign-in - IBM - Evan

Drinks: ahead of time

Food:

Cost: $2000 food/bev only

Registration: ticketbud

Promotion Channels:

- Austin AI Alliance Affiliated Event - https://www.austin-ai.org/events and newsletter / slack / linkedin
- IBM Austin contacts (Edward Calvsbert - VP Product et al, Mike Hollinger - DE Sustainability et al)
-

(day before) order food, confirm volunteers

Run of Show:

4:30PM organizers / volunteers meet at Q-branch downstairs

5PM - stage food

      Thundercloud party trays

5:30PM - 6:20PM Doors open & Mingling with drinks

6:20PM Welcome Remarks

6:30PM - 7PM Open Source Project Demos (mingling over food/bev)

- UT Machine Learning Lab Demo from https://www.linkedin.com/in/jwhorley/
- TX State Ventures Showcase - https://innovation.txst.edu/Programs-F/New-Ventures.html
- InstructLab Demo - IBM Austin Technical Vitality Council - Evan Blache or delegate
-

7:00PM - 7:15 PM Synthetic Data Panel (https://www.youtube.com/watch?v=vY7gmz3T8Kk)

7:15 PM - 7:30 PM Q&A

7:30 PM - 8:00 PM Networking

8:00 PM Doors Close

Panel Prep
Synthetic Data Teaser

Logistics:
15 min discussion
15 min Q&A

Purpose of the prep meeting:

- Everyone, meet everyone else!
  - How do you want to be introduced?
- Where're we going to agree, and disagree? Bring it on. :-) Let's have a debate.
- What questions should we address?
  - What're the challenging questions?
  - What do you think the audience'll want to know?
  - What're questions *you* can answer best?
  - What questions MUST you avoid?
  - What questions do you want to answer?
  - How do we have fun? Running jokes… secret word game? If we have fun, they'll have fun.

Panel Backgrounds:

Mike - Mike Hollinger is the president of the Austin AI Alliance and Chief Architect for Applied AI and Solutions in IBM Sustainability Software. An IBM Distinguished Engineer and Master Inventor, he has contributed to over 60 patents and invention disclosures.

Tahir - Professor of Analytics at McCoy College of Business at Texas State University-also serving as the director for TXST Center for Analytics and Data Science which houses the ExpandAI initiative.

Oji - Oji Udezue is a product exec who helped build recognizable companies and products at Microsoft, Atlassian, Twitter, Calendly, and Typeform. Now an Investor and international startup mentor. He is an AI product expert and consultant, Principal at ProductMind.

Jana - Tx State - new ventures - professor of innovation in Texas State mgmt dept - entrepreneur and investor (moderator)

Some topics / questions to choose from…

- Grounding - what's synthetic data in the first place?
  - What's the importance?
- Use cases?
  - InstructLab (aligning models) - IBM
  - Work order intelligence - GenAI + NLP - IBM
- Opportunities?
- Risks?
  - Amplifying imbalances in data
- Best practices
-

Impact of Simulated Data on AI and the Future (SXSW Submission)
Simulated data is revolutionizing AI by generating extensive, diverse datasets that propel innovation. This game-changing technology offers incredible opportunities to enhance AI development but also introduces significant risks that are often overlooked or external to the model. In this session, we'll delve into how simulated data is shaping the present and future of AI, examining its applications across various industries, the benefits it provides, and the challenges it presents. Discover how simulated data could redefine technology and impact multiple sectors.

Introduction (2 minutes)

Jana Moderator Opening:

- Welcome to this panel discussion "Simulated Data's Impact on AI and the Future."
- I'm Jana Minifie, I'm the director of TXST's Site for Entrepreneurship's Faculty Accelerator Innovation Program (FIAP) offering product-market fit training for TXST faculty and STEM-based startups from the community.
- Let me Introduce panelists with a quick highlight of their relevance to the discussion.
    - Mike - Mike Hollinger is the president of the Austin AI Alliance and Chief Architect for Applied AI and Solutions in IBM Sustainability Software. An IBM Distinguished Engineer and Master Inventor, he has contributed to over 60 patents and invention disclosures.
    - Oji: Oji Udezue is a product exec who helped build recognizable companies and products at Microsoft, Atlassian, Twitter, Calendly, and Typeform. Now an Investor and international startup mentor. He is an AI product expert and consultant, Principal at ProductMind.
    - Tahir: Professor at McCoy College of Business at Texas State University-also serving as the director for TXST Center for Analytics and Data Science which houses the ExpandAI initiative. Working with simulated data in fraud detection.
- Explain the flow: we will first have targeted questions to the panelists and audience, find common ground, and explore differences.
- During the network time, please meet with our panelists to learn more about them and about synthetic data.

Part 1: What's Synthetic Data and why's it important (6 minutes)

Focus on applications of simulated data, allowing panelists to establish expertise.

1. Let's define synthetic data. Tahir [to define in an engaging way] - example of why you should care and why its important. *This is a hook. Hyperbole helps… ;)*
2. Jana interjects , What's an example of a use case of synthetic data - Mike - from an Industrial perspective - can you share?
    a. ChatGPT only works because of synthetic data… derived from Gartner
3. Oji, How about from a product perspective? If you're a small company, how does synthetic data help with decision-making?
    a. Frame it bigger… synthetic humans, or behavior… UI decisions, unit testing, load tests, performance testing (even before genAI)
        i. Follow-on at some point - Mike - GenAI use case that is only possible because of GenAI and synthetic data. Because we don't want people having to label ~10,000 items. We only want them to label 50.

Part 2: Best Practices (6 minutes)

Compare benefits and risks to identify agreements.

1. **To Oji** (Product Management):
   "Oji, product teams are often tasked with balancing speed and reliability. How does simulated data help PMs test features or improve user experiences without waiting for real-world data?"
2. **To the Panel (Moderator-led Pivot):**
   "Everyone seems to agree simulated data speeds up AI development. But what's the trade-off? For instance, does the reliance on simulated data introduce risks, like overfitting to non-realistic scenarios?"
   (Encourage panelists to comment and build on one another's points.)

(adjust questions here..)

Part 3: Debating the Future (5–6 minutes)

Highlight disagreements or contrasting views on trends and ethics.

1. **To Mike Hollinger or Oji Udezue**:
   "Mike/Oji, some argue that as simulation fidelity improves, it could fully replace real-world data collection in certain areas. Do you see that happening, or is there always a need for real-world grounding?"
2. **To the Panel (Ethics Question):**
   What's the effect of synthetic data on the future of datasets, and AI, looking forward to 2030 or beyond? When does the AI snake start eating its tail?
3. ***Transparency around use of data***
   a. *Its about **Trust.** Open Source models, open source data, open source pre-training / aligning, open weights… all drive a better understanding / trust on the system.*

Mike - grab homework from IBV synthetic data / ethics

Closing Question (1–2 minutes)

To the Panel:
*"In one sentence: What's the way to take responsibility and drive an ethics-centered approach to **partner** with AI more - how do we keep humans in the loop?" (create a hopeful end)*