rnl national conference

leading ai innovation

empowering higher education

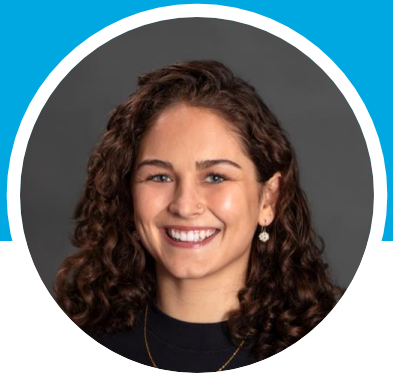
How to Keep Your Digital Assistant from Making Headlines for the Wrong Reasons

Friederike Maag – AI Solutions Consultant, RNL

Agenda

1. Introductions
2. High-level Digital Assistant Overview
3. Examples of Digital Assistant Fails
4. Key Takeaways
5. Questions & Next Steps

Introductions



Friederike Maag

AI Solutions Consultant

3+ years of experience in conversational and generative AI as lead solution architect and product manager at OneReach.ai, MS in Business and Social Impact and BS in Neuroscience from USC

Ashwin Kannan

Conversational AI Developer

5+ years of experience developing machine learning and AI solutions at MuSigma Inc., Tata Consultancy Services, and Uniphore. MS in Electrical and Computer Engineering from Arizona State University.

High-Level Digital Assistant Overview

What is a Digital Assistant?

A computer program that **simulates human conversation** with an end user.

- Not all are equipped with artificial intelligence (AI)
- Sophistication is dependent on technology stack, data availability/quality, and user experience design
- Use cases include: communicating with users to answer questions, self-service tasks, and transactional automations

Pretty simple, right? What could possibly go wrong?

Pretty simple, right? What could possibly go wrong?

A lot.

Examples of Digital Assistant Fails

DATA

Twitter taught Microsoft's AI chatbot to be a racist in less than a day

Pitfall: Poor-quality, un-monitored data.

Consequence: Unintended amplification of societal prejudices.

Source: https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist

TESTING

Air Canada chatbot promised a discount. Now the airline has to pay it.

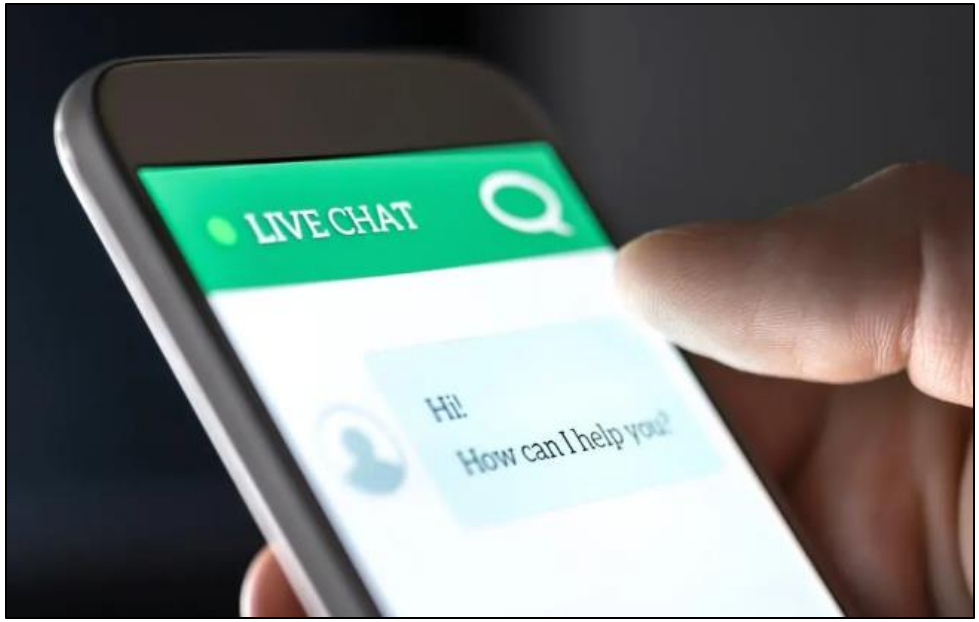
Pitfall: Absence of accuracy testing.

Consequence: Distribution of misinformation leading to reputational degradation.

Source: https://www.washingtonpost.com/travel/2024/02/18/air-canada-airline-chatbot-ruling/

SECURITY

Businesses Warned Over Risks Of Chatbot Prompt Injection Attacks



Pitfall: Training data tampering and lack of input validation.

Consequence: Potential system security breach.

Source: https://www.forbes.com/sites/emmawoollacott/2023/08/30/businesses-warned-over-risks-of-chatbot-prompt-injection-attacks/?sh=4b7bdb035b03

California Law Bans Bots From Pretending to Be Human

SECURITY

Pitfall: Impersonation of human and lack of clear and conspicuous disclosure.

Consequence: Fraud, potential legal repercussions, fines, etc.

Sources:

- https://www.pcmag.com/news/california-law-bans-bots-from-pretending-to-be-human
- https://www.newyorker.com/tech/annals-of-technology/will-californias-new-bot-law-strengthen-democracy

How Strangers Got My Email Address From ChatGPT's Model

SECURITY

Pitfall: Absence of data privacy safeguards.

Consequence: Sensitive data leaks to unintended destinations.

Source: https://www.nytimes.com/interactive/2023/12/22/technology/openai-chatgpt-privacy-exploit.html

Key Takeaways

4

Key Takeaways

How to mitigate risk and avoid pitfalls

DATA

Train models with high-quality data

TESTING

Test rigorously before and after deployment

ITERATION

Continue to refine based on interaction data and user feedback

SAFETY

Keep system security, data privacy, and regulatory compliance top of mind

Key Takeaways

How to mitigate risk and avoid pitfalls

DATA

Train models with high-quality data

TESTING

Test rigorously before and after deployment

ITERATION

Continue to refine based on interaction data and user feedback

SAFETY

Keep system security, data privacy, and regulatory compliance top of mind

Key Takeaways

How to mitigate risk and avoid pitfalls

DATA

Train models with high-quality data

TESTING

Test rigorously before and after deployment

ITERATION

Continue to refine based on interaction data and user feedback

SAFETY

Keep system security, data privacy, and regulatory compliance top of mind

Key Takeaways

How to mitigate risk and avoid pitfalls

DATA

Train models with high-quality data

TESTING

Test rigorously before and after deployment

ITERATION

Continue to refine based on interaction data and user feedback

SAFETY

Keep system security, data privacy, and regulatory compliance top of mind

rnl compass

The digital assistant for higher education

Product overview

- ✓ **Real-time communication with curated and validated content:**
  Connect with today's students, parents, and alumni through authentic, conversational interactions.
- ✓ **24/7/365 availability:**
  Ensure 24/7 effective engagement, leading to more applications.
- ✓ **CX improvement:**
  Improve student engagement and satisfaction with a human-like and empathetic experience.
- ✓ **Improve employee satisfaction while reducing costs:**
  Save valuable staff time with automated assistance and support, freeing up time for other key deliverables and projects. Enhance the enrollment experience with seamless interactions.
- ✓ **CRM and website integration:**
  Bi-directional and real-time connections to your existing student data.
- ✓ **100% secure data:**
  Ensuring responses are based on only university information.

Questions & Next Steps

rnl national conference

leading ai innovation

empowering higher education

Dallas, Texas | July 23-25, 2024