



Assessing the Impact of AI-Based Meeting Summarization on Information Relevance & Engagement

Jimmy Spoerl, Daniel Choi, Jeremy Locklear, Henry Engler, and Zachary Leith

Advisor: LTC Dan Koban

United States Military Academy

April 23rd, 2026

Experimental Setup

- 8x Podcast summaries
- Human vs. AI-generated



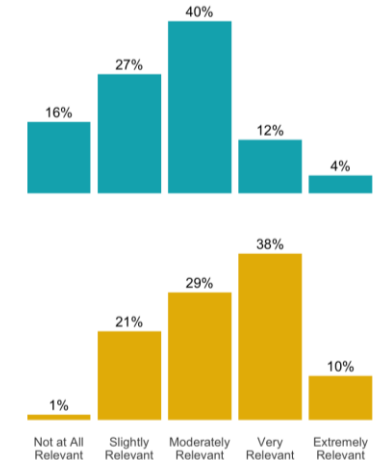
Evaluation (N = 304 cadets)

- Perceived relevance
- Listening Intentions



Results

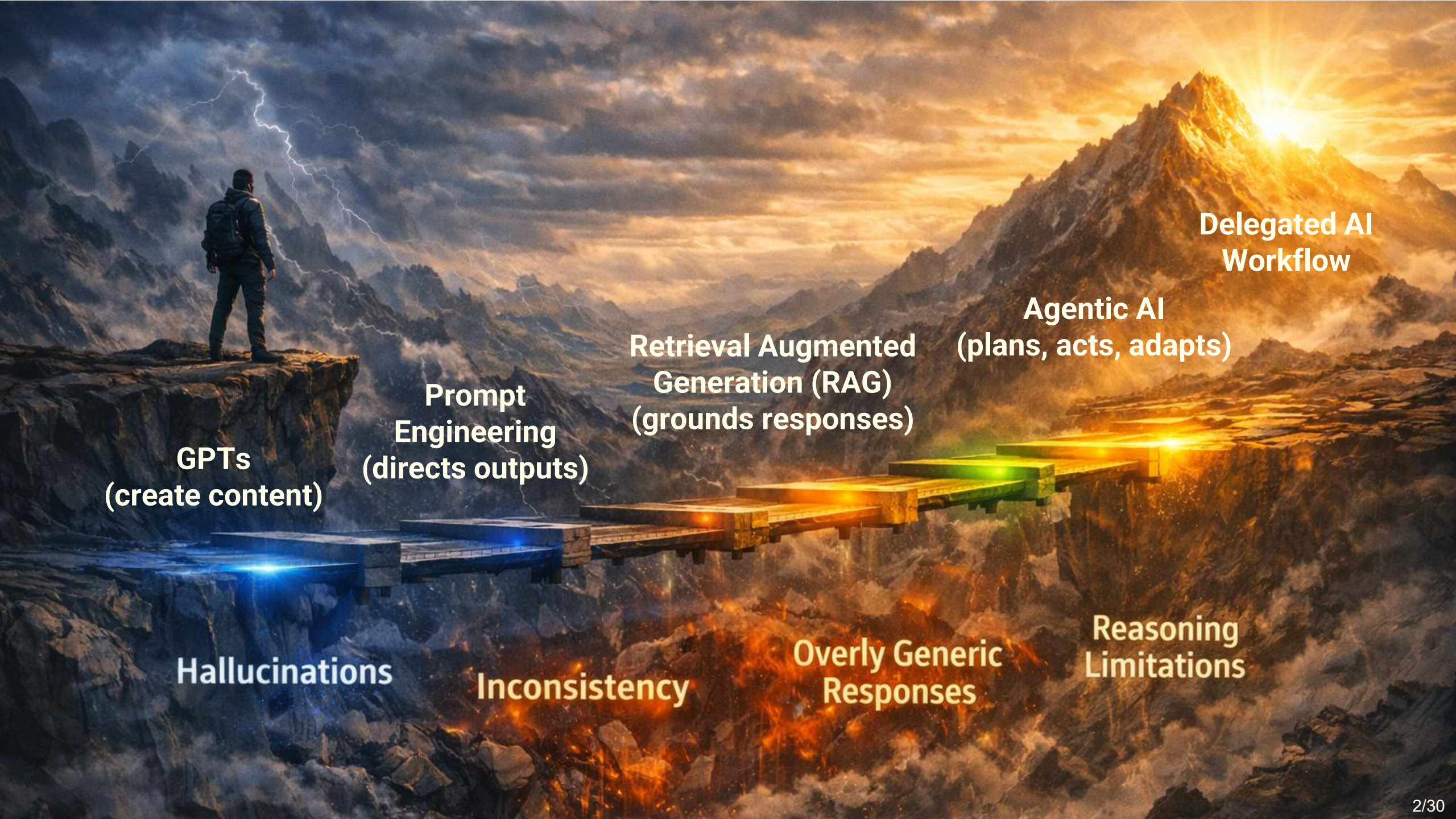
- AI summaries rated more relevant
- AI increased listening intentions



The views expressed herein are those of the authors and do not reflect the position of the United States Military Academy, the Department of the Army, or the Department of Defense.

Problem Background

- Daily update briefs are critical for situational awareness and decision-making. However, competing work obligations often limit attendance.
- Manually summarizing meetings is time consuming and prone to error.



**Delegated AI
Workflow**

**Agentic AI
(plans, acts, adapts)**

**Retrieval Augmented
Generation (RAG)
(grounds responses)**

**Prompt
Engineering
(directs outputs)**

**GPTs
(create content)**

**Reasoning
Limitations**

**Overly Generic
Responses**

Inconsistency

Hallucinations

Related Work

- Personalized summarization increased summary relevance by 9%.¹
- However, performance was evaluated **using a GPT4-based metric** (AUTOCALIBRATE).²

GAP: The previous study didn't evaluate how *people* view and engage with AI-generated summaries

¹ Kirstein et al. (2024). *Tell me what I need to know: Exploring LLM-based (Personalized) Abstractive Multi-Source Meeting Summarization*. arXiv. <https://doi.org/10.48550/arxiv.2410.14545>

² Liu et al. (2023). *Calibrating LLM-based evaluator*. arXiv. <https://doi.org/10.48550/arxiv.2309.13308>

Hypothesis: People will prefer AI-generated summaries over the human-generated summaries.

Experimental Setup

- 8x Podcast summaries
- Human vs. AI-generated



Evaluation (N = 304 cadets)

- Perceived relevance
- Listening Intentions



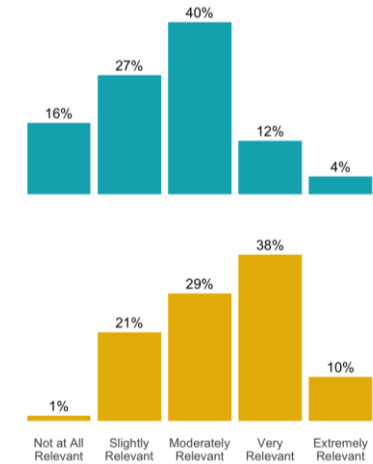
Human



Copilot

Results

- AI summaries rated more relevant
- AI increased listening intentions



Design Choice #1:

What should we use as a proxy for classified update briefs?



Monthly Community Leader Information Forums



Congressional Testimonies

Design Choice #1:

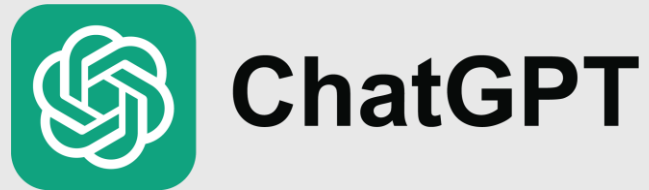
What should we use as a proxy for classified update briefs?



Topic	Episode	Title
Leadership	L1	Leadership Lessons on Building Teams and Champions (Coach Krzyzewski)
	L2	Retaining Good Talent Through Inspiring Leadership (COL Everett Spain)
Data	D1	Decoding Data Literacy (COL Nicholas Clark)
	D2	Bridging Human and Machine through Data Science (COL David Beskow)
Technology	T1	Exploring Space and Modern Warfare (LTC Will Koch)
	T2	Leadership & Technology on the Battlefield (LTG Scott Miller)
Strategy	S1	Global Security (NATO Secretary General Mark Rutte)
	S2	Building Cross-Cultural Competence Through Language and Cultural Immersion (Colonel John Baskerville)

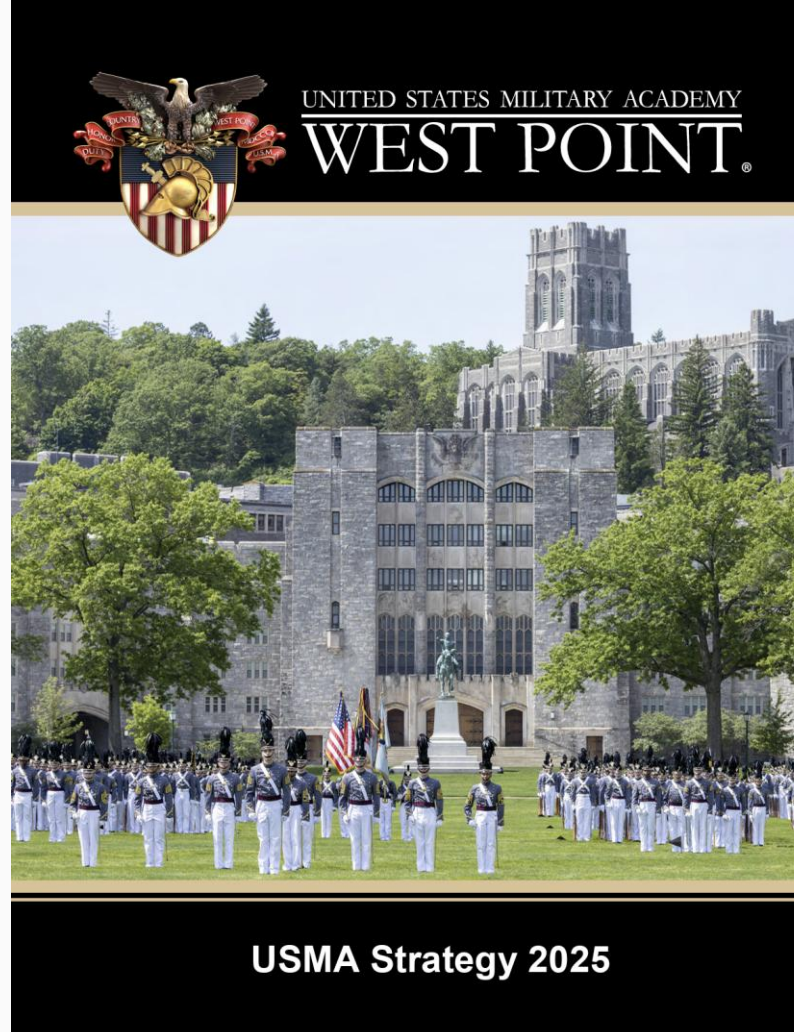
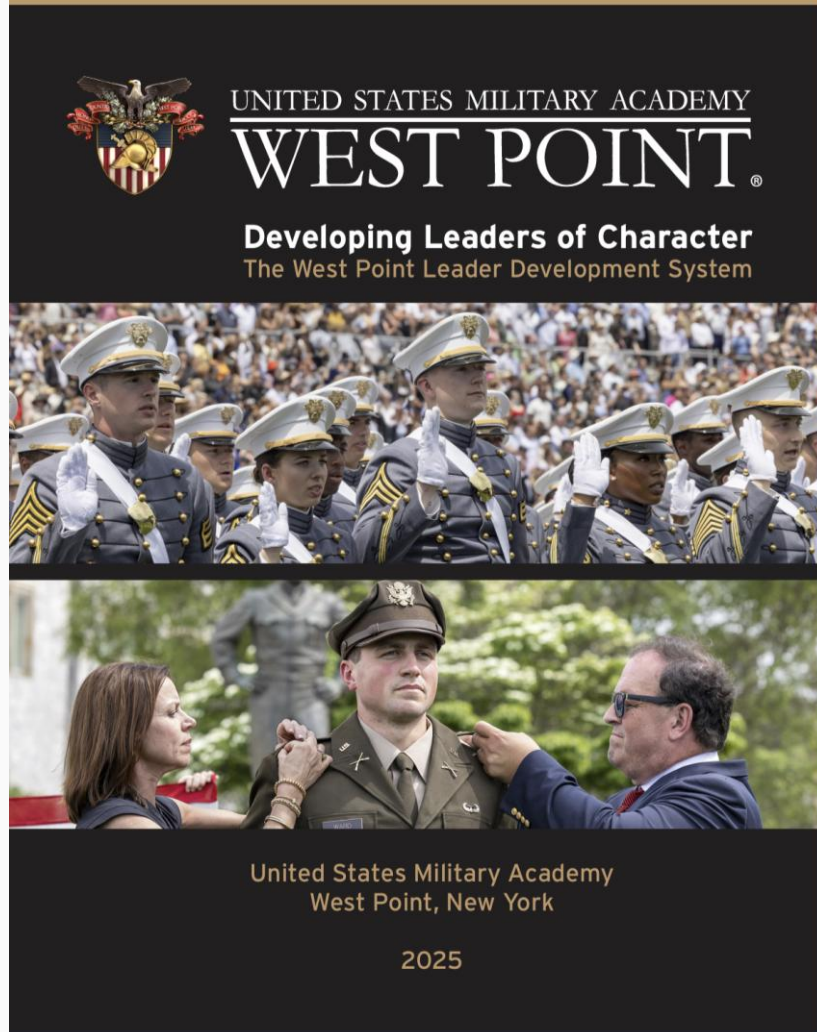
Design Choice #2:

Which model to use? There are tons to choose from...



Design Choice #3:

What documents should we use for reference texts?



Academic Program Goals (APGs)

What Graduates Can Do Statements (WGCDs)

The Overarching Academic Goal: Graduates integrate knowledge and skills from a variety of disciplines to anticipate and respond appropriately to opportunities and challenges in a changing world.

APG 1. Communication: Graduates communicate effectively with all audiences.

- WGCD 1.1: Listen actively, read critically, and develop an informed understanding of the communications of others.
- WGCD 1.2: Speak and write using Standard American English.
- WGCD 1.3: Effectively convey meaningful information to diverse audiences using appropriate forms and media.
- WGCD 1.4: Communicate in a foreign language.
- WGCD 1.5: Use sound logic and relevant evidence to make convincing arguments.

APG 2. Critical Thinking and Creativity: Graduates think critically and creatively.

- WGCD 2.1: Identify the essential aspects of a situation and ask relevant questions.
- WGCD 2.2: Integrate knowledge and skills from a variety of disciplines.
- WGCD 2.3: Make meaningful connections and distinctions among diverse experiences and concepts.
- WGCD 2.4: Reason both quantitatively and qualitatively.
- WGCD 2.5: Think innovatively and accept risk to pursue solutions in the face of ambiguity.
- WGCD 2.6: Value reflection and creativity; envision possibilities.

APG 3. Lifelong Learning: Graduates demonstrate the capability and desire to pursue progressive and continued intellectual development.

- WGCD 3.1: Demonstrate the willingness and ability to learn independently.
- WGCD 3.2: Engage successfully in deliberate self-directed and collaborative learning experiences.
- WGCD 3.3: Pursue self-awareness and embrace the responsibility for personal intellectual development.
- WGCD 3.4: Pursue knowledge in areas of personal or professional interest.

APG 4. Ethical Reasoning: Graduates recognize ethical issues and apply ethical perspectives and concepts in decision making.

- WGCD 4.1: Understand the intellectual foundations of ethical principles.
- WGCD 4.2: Recognize ethical components of problems and situations.
- WGCD 4.3: Examine and evaluate different ethical perspectives, principles, and concepts in context.
- WGCD 4.4: Apply ethical perspectives and concepts in solving complex problems, including those found in military settings.

Design Choice #4:

What instructions should we give to our CoPilot?

Agent Builder > Podcast Summarizer

Last updated December 3, 2025 at 3:57:17 PM [Update](#) [Share](#) ...

[Describe](#) [Configure](#)

Details

Name

Podcast Summarizer

Description ⓘ

Summarize podcasts

Instructions

 ⓘ

Agent preview



Podcast Summarizer

Summarize podcasts

Type your message 

Knowledge

 ⓘ

Choose the sources your agent will use to generate responses

Add specific websites ⓘ

Search all websites. To limit results to specific sites, [enter a URL above](#) ⓘ ☐

Only use specified sources ⓘ ☐

**Basic users can't use
Copilot's knowledge base**

Mission: Given a podcast transcript, **produce a concise, high-impact leadership summary designed for Army leaders**, using a format optimized for persuasion, clarity, and retention.

CoPilot Instructions

- 50-word summary
- Three key insights
- Bottom line

Content Rules

- Use transcript content only
- Never invent facts, quotes or context
- Factual grounding

Study Conditions

DECODING DATA LITERACY: PERSPECTIVES FROM COLONEL NICHOLAS CLARK

Published on: 4th September, 2023

Join us for another episode of *Inside West Point: Ideas That Impact*. In this episode, Colonel Nicholas Clark and West Point Dean Brigadier General Shane Reeves explore the pivotal role of data-driven decisions in the military and. **Clark sheds light on the inception of the Center for Data Analysis and Statistics, his creation of a data literacy program** for the Army, and the surging significance of data and analytics. Clark underscores the imperative for officers to possess data literacy and unveils his research award focused on engaging undergraduates in data science. Tune in to discover how to enhance your data literacy and gain insights into the fulfilling realm of teaching at West Point.

Status Quo:
High-level
wall of text



Study Conditions

DECODING DATA LITERACY: PERSPECTIVES FROM COLONEL NICHOLAS CLARK

Published on: 4th September, 2023

Afghanistan's withdrawal exposed a silent threat: **leaders drowning in messy, unstructured data**. Colonel Nicholas Clark warns that without "data-informed" judgment, **commanders risk paralysis at the point of decision**. Officers who master data literacy—asking the right questions, shaping raw inputs, and communicating insights—will weaponize information for speed, clarity, and mission success.

Three Key Insights:

- **Look for chaotic "data deserts"** and overwhelming raw feeds that obscure operational decisions.
- Strive for disciplined data literacy—**ask sharper questions**, format inputs, and enforce analytics in briefs.
- **Gain faster decisions and readiness** by turning sensors into solutions at the edge of combat.

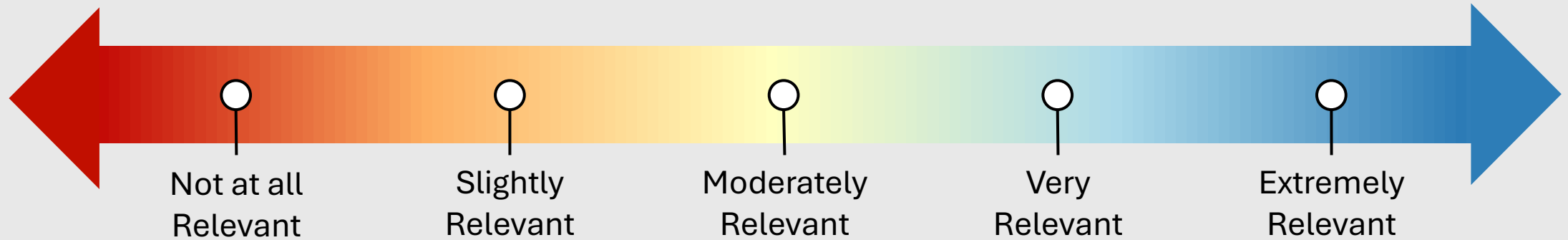
Listen to Colonel Clark explain why data literacy is now a warfighting imperative for every Army leader.

CoPilot:
Meaningful,
structured excerpts

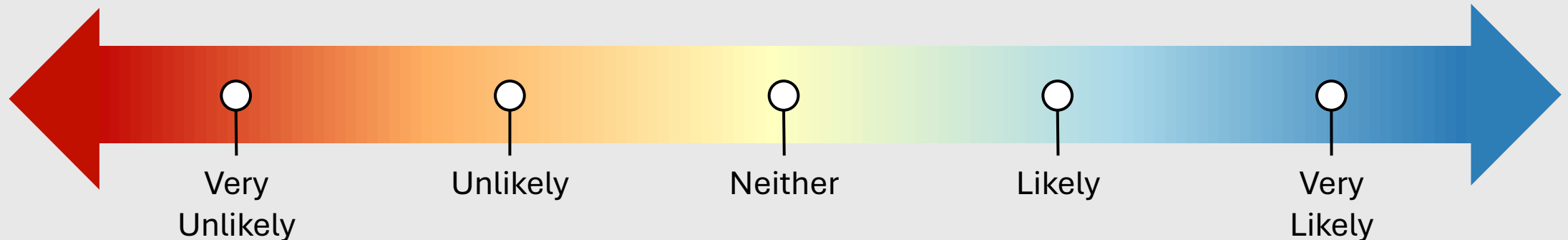


Primary Outcomes

1. How relevant does this episode feel to you personally?



2. How likely would you be to listen to this episode if you saw this summary?



Design Choice #5: How to structure and administer the survey

Info & Consent

12:29

***Research Study
Information and Consent**

What You Will Do
If you choose to participate, you will:

1. Read **five short podcast summaries** on topics relevant to West Point.
2. Answer **three brief questions** about each summary:
 - how relevant it feels,
 - how valuable it seems, and
 - how likely you are to listen to the episode.
3. Complete a few short background questions.

The survey will take **approximately 10–15 minutes**. There are no right or wrong answers—just respond honestly

Episode Summary

12:29

***Episode 1
Winning Culture with Coach K:
Leadership Lessons on
Building Teams and Champions**

Under relentless pressure, leaders either adapt or break. Coach K reveals that winning isn't about talent alone—it's about preparation, trust, and adaptability. From West Point cadet to Olympic champion, his lessons demand leaders embrace failure, improvise under uncertainty, and build "speed to trust"—because warfighting and leadership share the same unforgiving terrain.

Three Key Insights:

- Look for environments that force failure—"failure is not our destination"—to build resilience and agility.
- Strive for adaptability and honest communication:

Questions

12:29

1. How relevant does this episode feel to you personally?

☐ Not at all relevant

☐ Slightly relevant

☐ Moderately relevant

☐ Very relevant

☐ Extremely relevant

***2. How valuable do you think this episode would be for your development (as a leader / cadet)?**

☐ Not at all valuable

☐ Slightly valuable

☐ Moderately valuable

☐ Very valuable

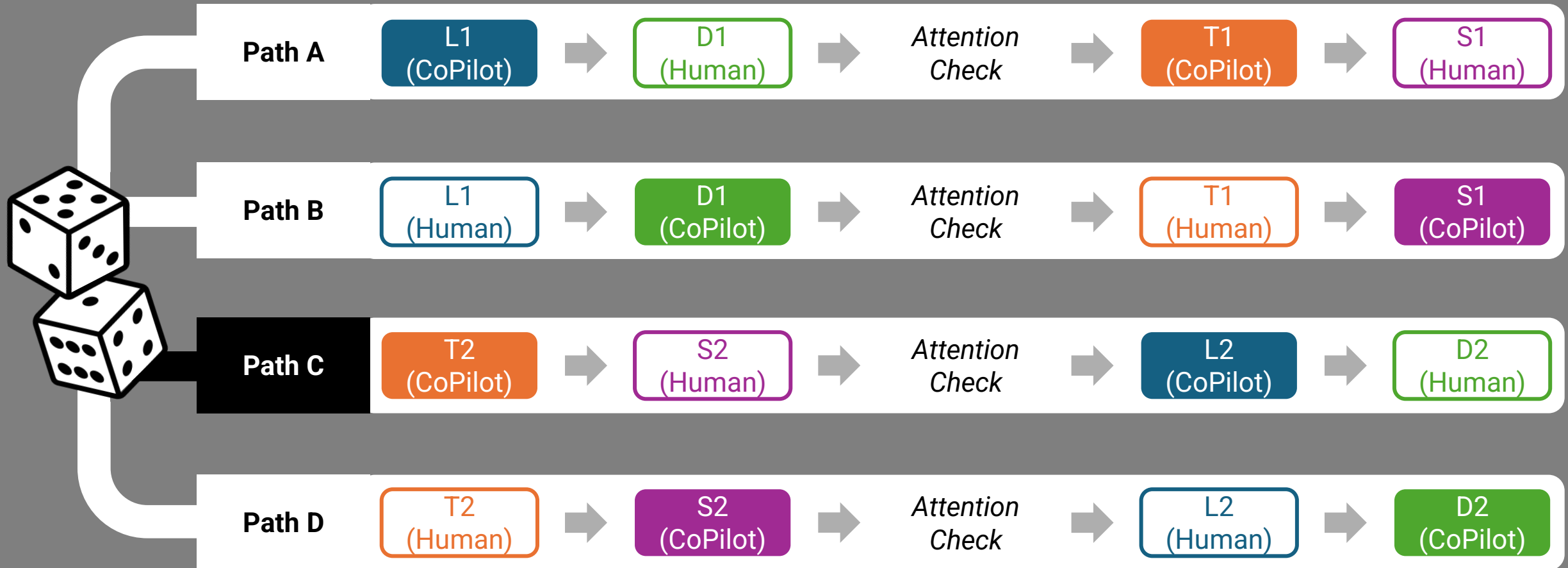
Attention Check

12:29

***Episode 3
Reading Instructions and
Relevance Judgments**

In this special episode, the host talks with a behavioral scientist about why people often skim instructions and how that can affect the quality of their decisions. They explain that, in research studies, one way to check whether participants are paying attention is to include a simple instruction in the summary and see if people follow it. For the purposes of this survey, we are using this episode as a quality-check item. **If you are reading this summary carefully, please select "Moderately relevant" for every question that appears below, even if you would normally choose something else.** This helps the research team verify that responses are thoughtful and that people are

Design Choice #6: Develop a Randomization Strategy



Design Choice #7: How can we recruit people to take our survey?



[Home](#) [My Studies](#) [All Studies](#) [Add New Study](#) [Prescreen Results](#) [My Profile](#) [Logout](#)

Koban Dan (Researcher)

[All](#) [Active](#) [Inactive](#)

Sort by: [Study Name](#) [Last Activity Date](#)

Study Information	Status	View
A Comparison of the Combat Application Tourniquet and Smart Tourniquet (10 Points) The Smart Tourniquet is an interactive training tool used for combat life-saving measures. Participants will interact with a manikin in a training environment with the intent to improve overall combat application tourniquet training..	Approved Timeslots Available	- Study Info - Timeslots
AI Adoption and Ethical Use (10 Points) (Online Study) This study examines cadet beliefs about artificial intelligence (AI) use, familiarity, attitudes, and performance over time, with emphasis on their connection to moral reasoning and ethical leadership in a military academic setting.	Approved No Available Timeslots	- Study Info - Timeslots
AI Chatbot (15 Points) Participants in this study will complete an online quiz with the assistance of an AI chatbot, as well as complete surveys and cognitive tasks, all of which can be done remotely using a desktop or laptop computer.	Approved No Available Timeslots	- Study Info - Timeslots
AI Chatbot Study (15 Points) (Online Study) Participants in this study will complete an online quiz with the assistance of an AI chatbot, as well as fill out surveys and tests, all of which can be done remotely using a desktop or laptop computer.	Approved No Available Timeslots	- Study Info - Timeslots

Study Participants

N = 335

Failed Attention Checks

N = 31

Final Used in Analysis

N = 304

Demographics



Class Composition

49% Upper-Class

51% Lower-Class



Academic Background

55% STEM

45% Non-STEM



Gender

72% Male

27% Female



Prior Podcast Exposure

66% No Prior Exposure

34% Some Prior Exposure



Race / Ethnicity

66% White

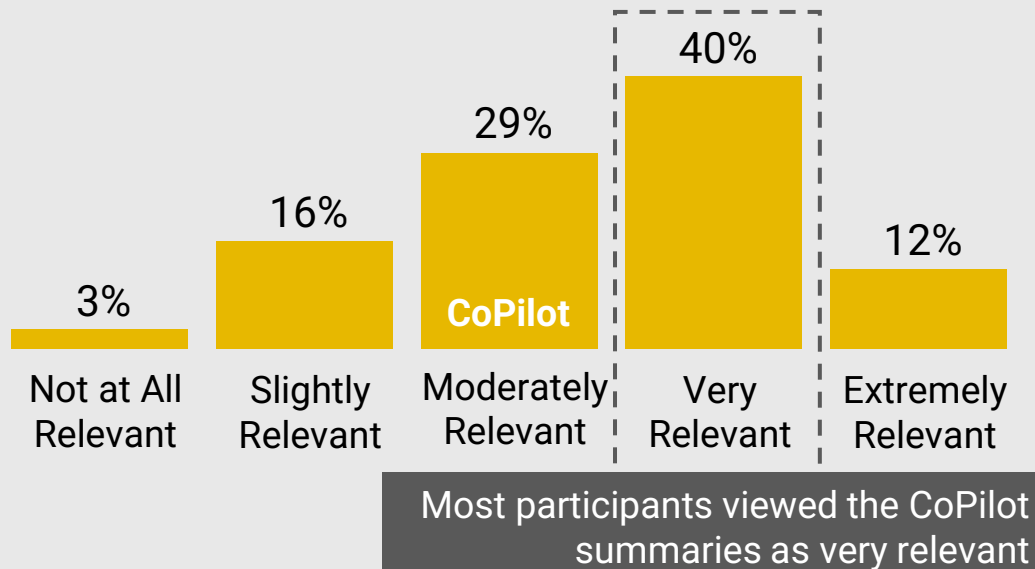
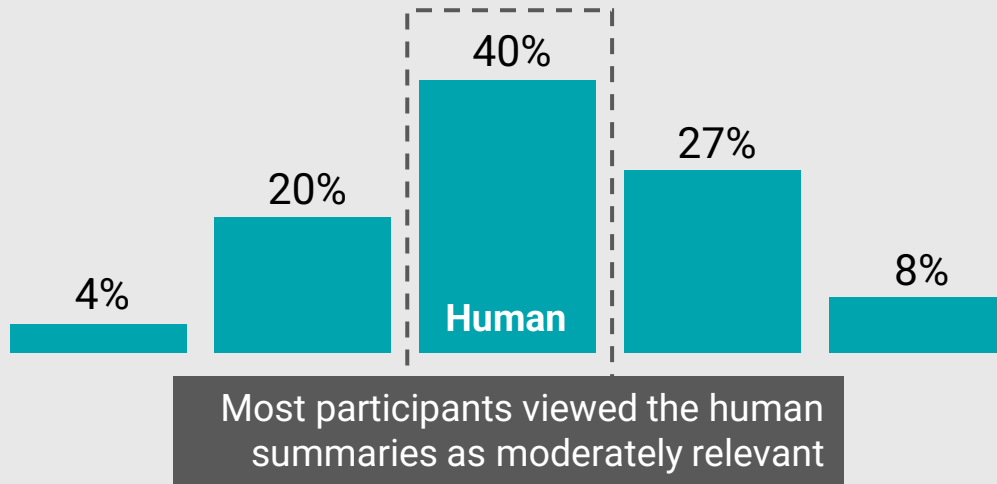
17% Asian

6% Black

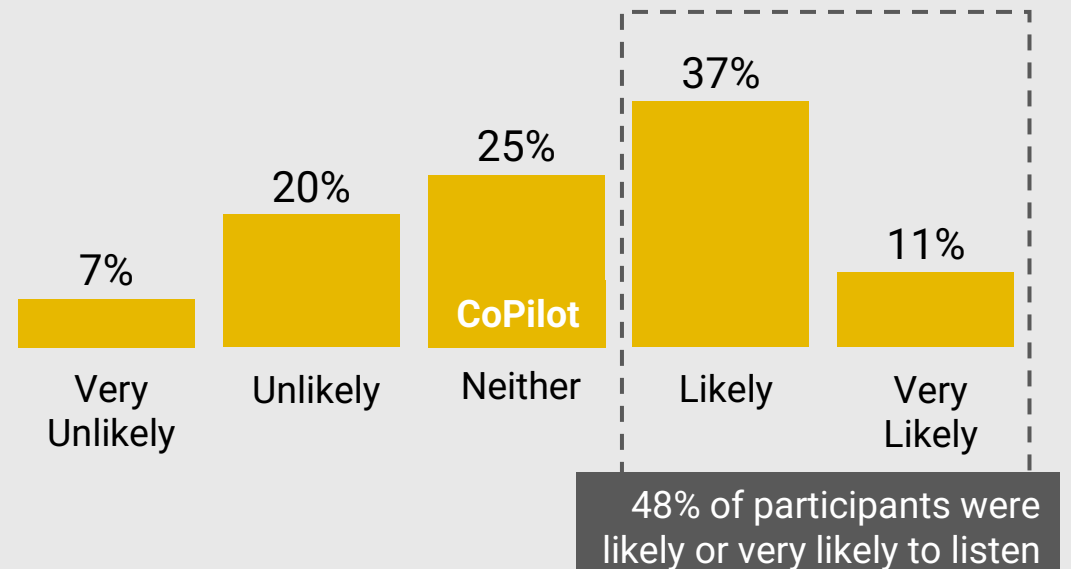
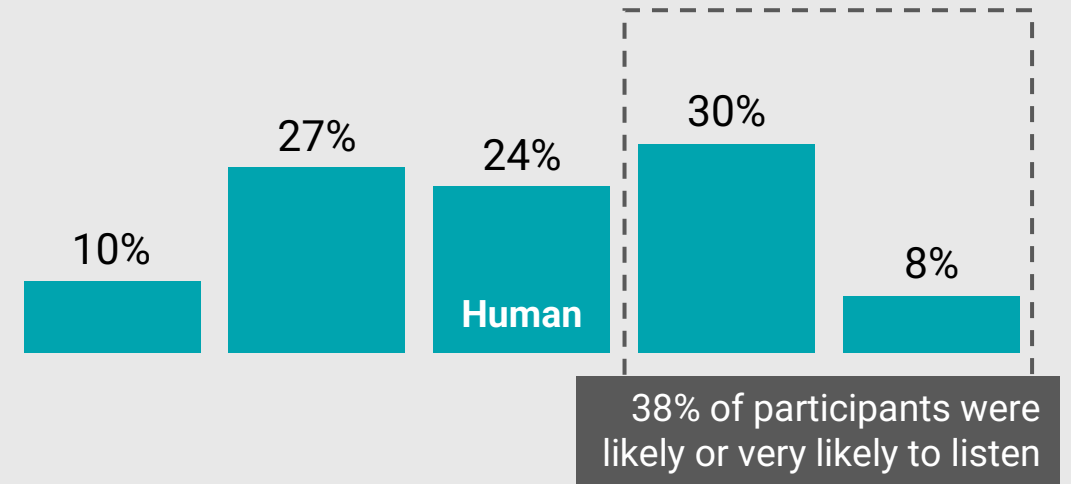
5% Hispanic

6% Other / Prefer not to say

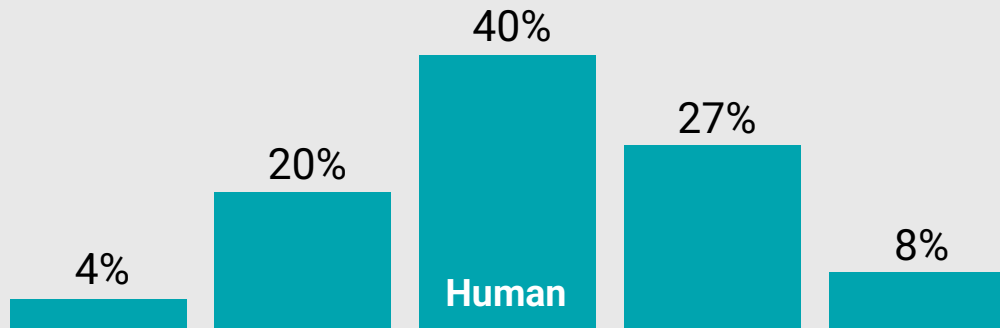
Relevance



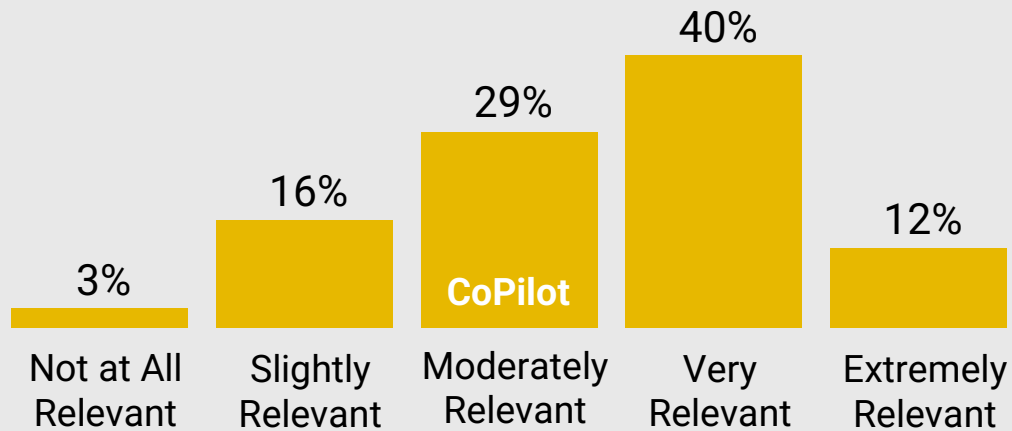
Intentions to Listen



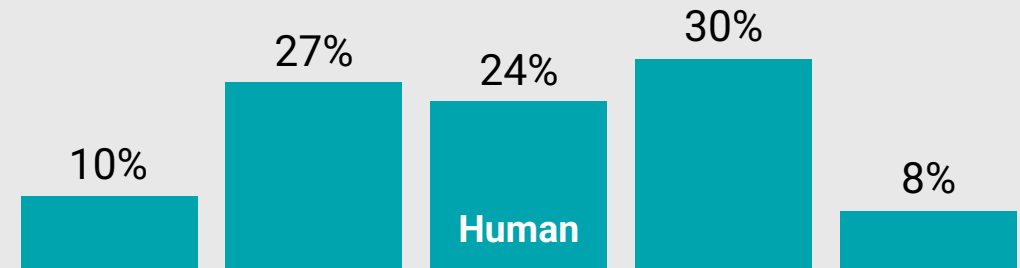
Relevance



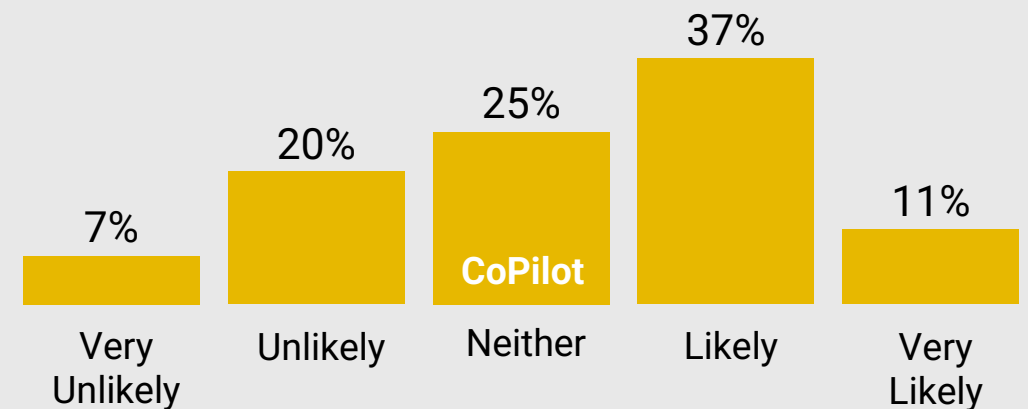
Cliff's δ (Ordinal effect size)
AI summaries rated higher than human summaries
16% (95% CI = [10%, 22%]) more often than the reverse



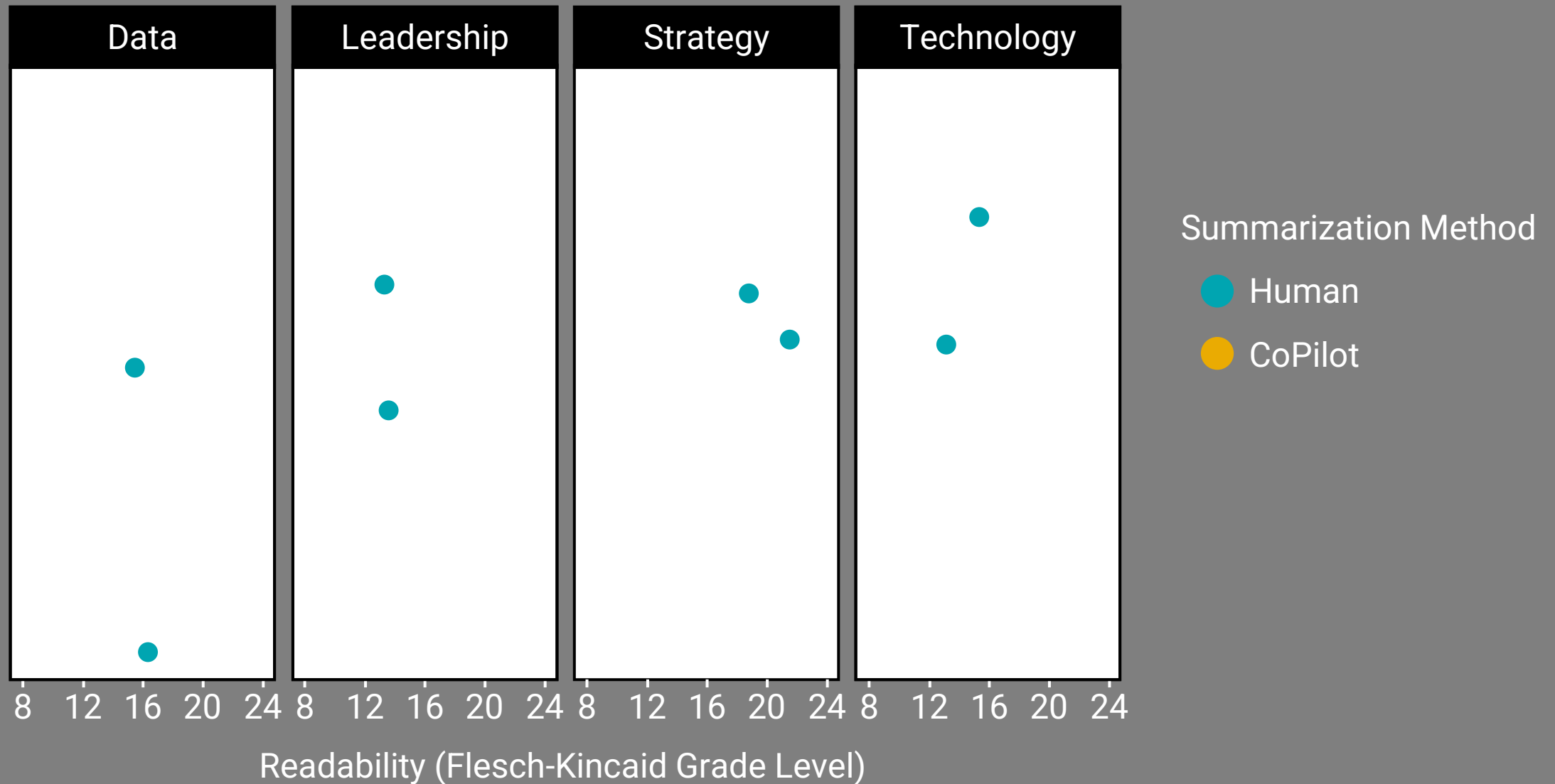
Intentions to Listen



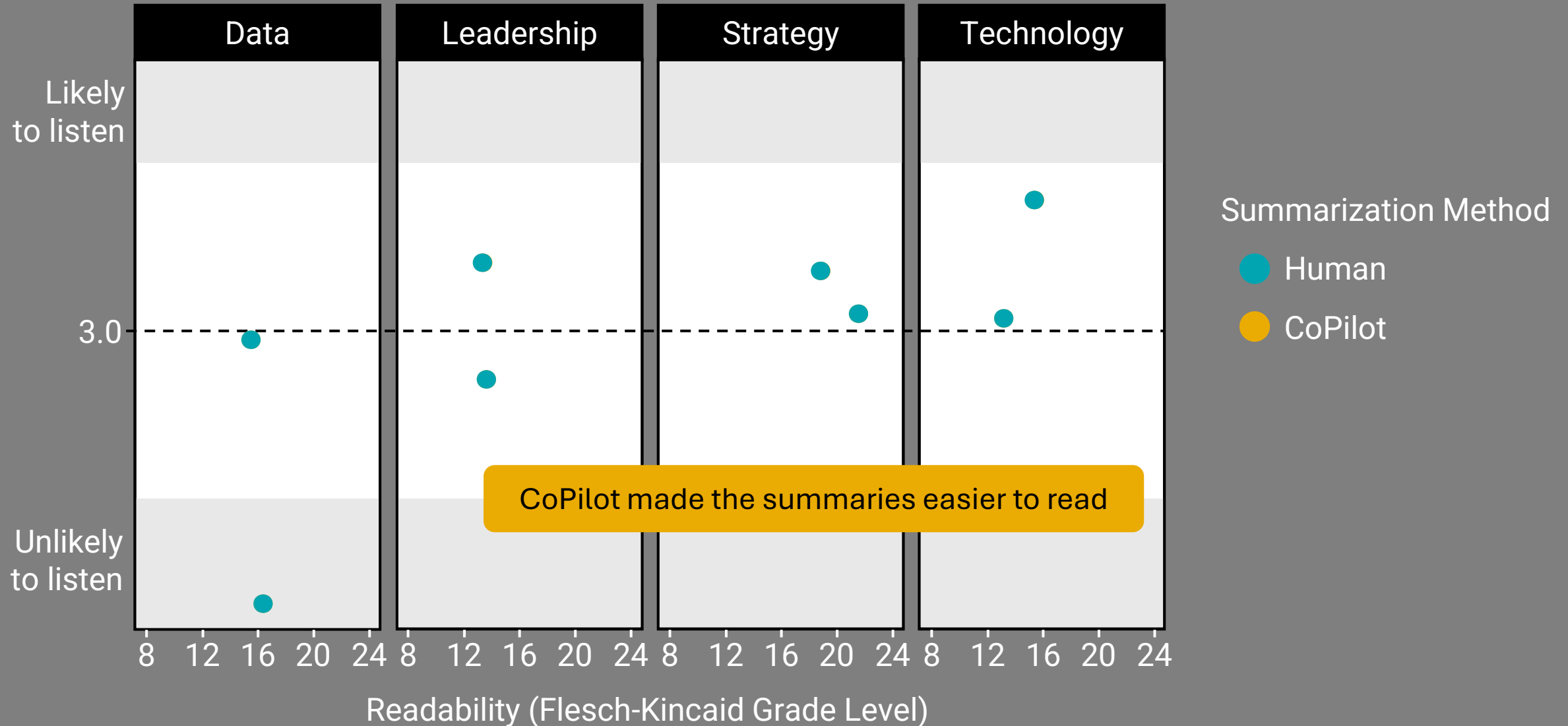
Cliff's δ (Ordinal effect size)
AI summaries rated higher than human summaries
12% (95% CI = [6%, 19%]) more often than the reverse



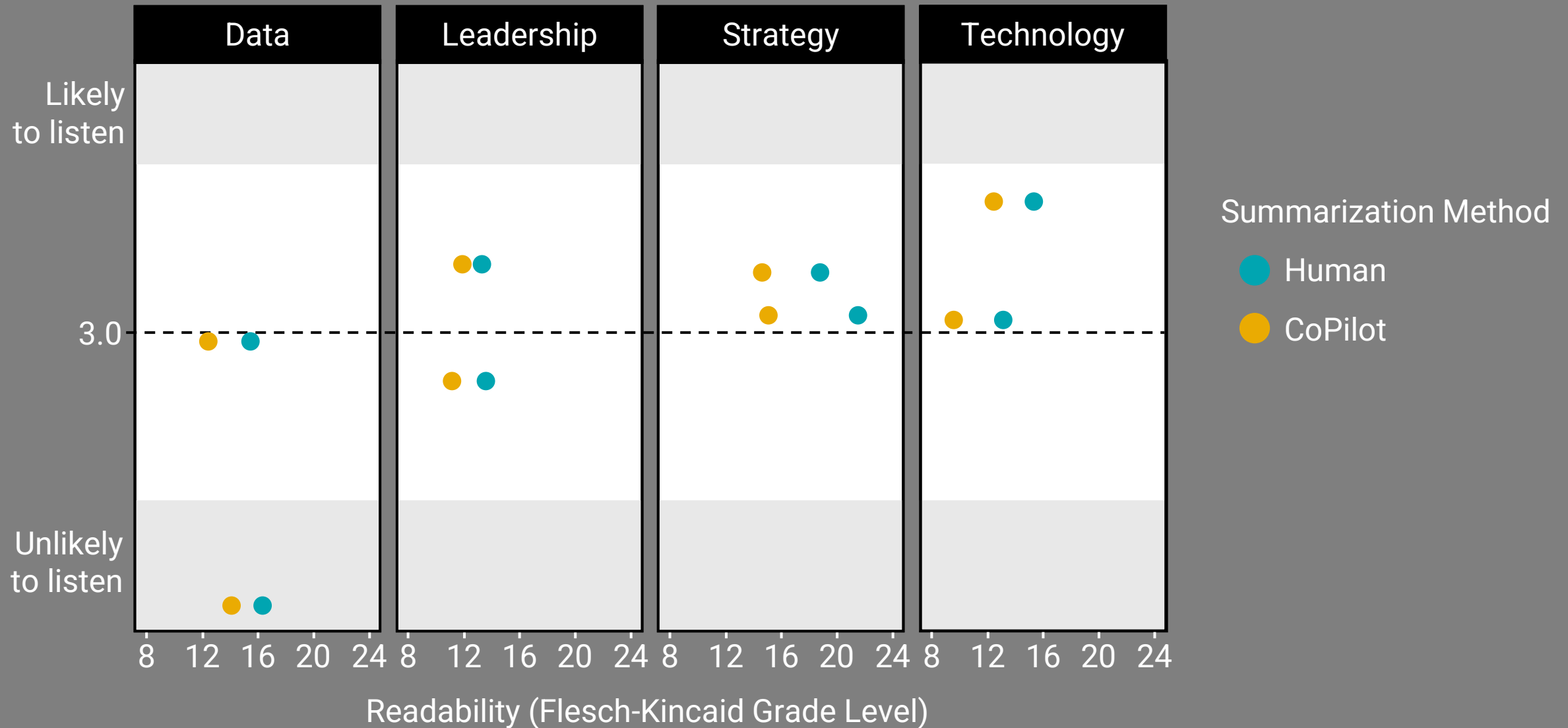
Intentions to Listen



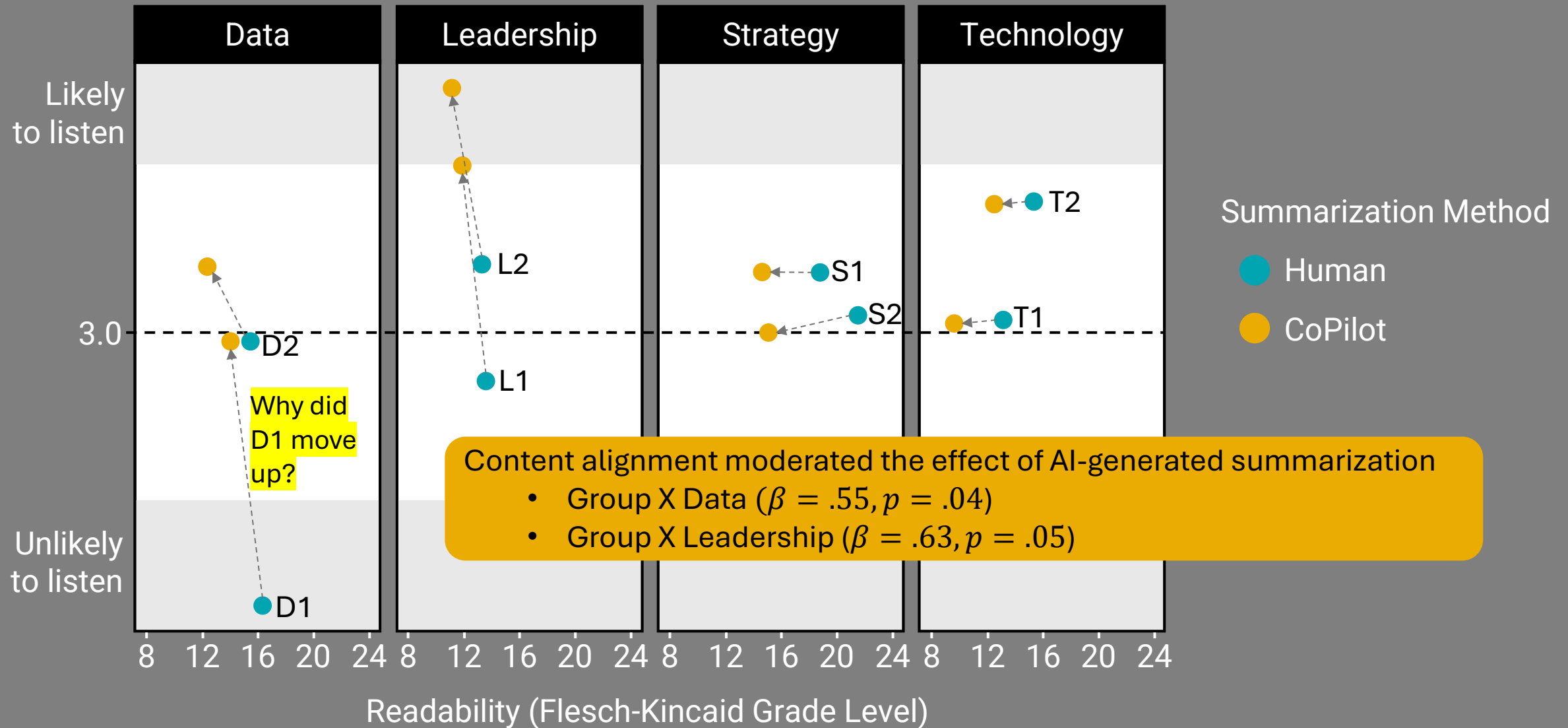
Intentions to Listen



Intentions to Listen



Intentions to Listen

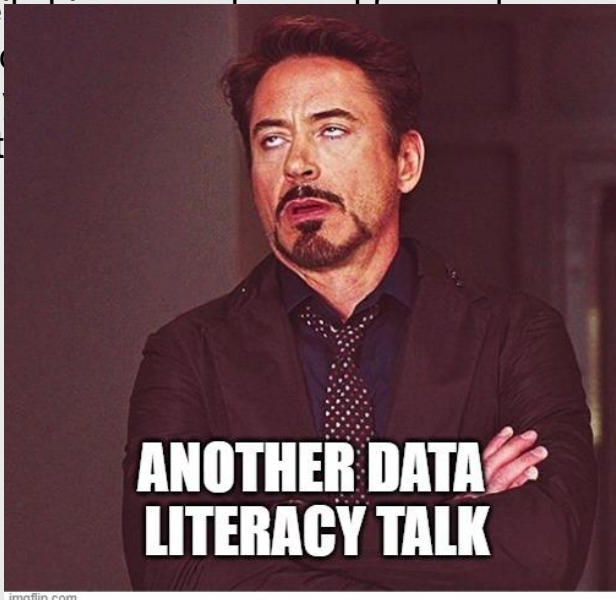


D1: Decoding Data Literacy: Perspectives from Colonel Nicholas Clark

Control: Human-generated

Podcast Description

Join us for another episode of *Inside West Point: Ideas That Impact*. In this episode, Colonel Nicholas Clark and West Point Dean Brigadier General Shane Reeves explore the pivotal role of data-driven decisions in the military and. **Clark sheds light on the inception of the Center for Data Analysis and Statistics, his creation of a data literacy program for the Army**, and the surging significance of data and analytics. Clark underscores the imperative for officers to possess data literacy and unveils his insights on managing undergr... cover how to enhance... to the fulfilling realm of t



Treatment: GPT-generated

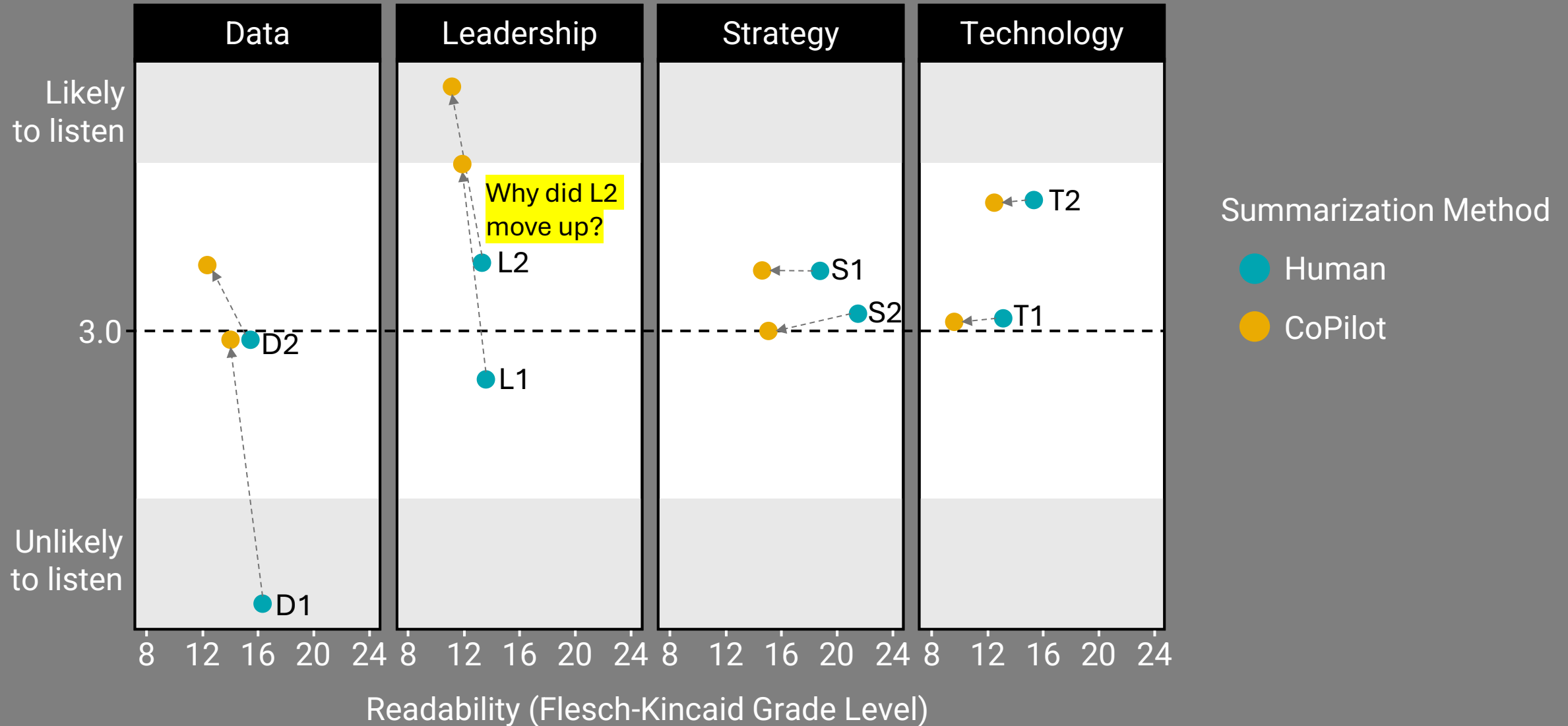
Afghanistan's withdrawal exposed a silent threat: **leaders drowning in messy, unstructured data**. Colonel Nicholas Clark warns that without “data-informed” judgment, commanders risk paralysis at the point of decision. Officers who master data literacy—asking the right questions, shaping raw inputs, and communicating insights—will weaponize information for speed, clarity, and mission success.

Three Key Insights:

- **Look for chaotic “data deserts”** and overwhelming raw feeds that obscure operational decisions.
- Strive for disciplined data literacy—**ask sharper questions**, format inputs, and enforce analytics in briefs.
- **Gain faster decisions and readiness** by turning sensors into solutions at the edge of combat.

Listen to Colonel Clark explain why data literacy is now a warfighting imperative for every Army leader.

Intentions to Listen



L2: Retaining Good Talent Through Inspiring Leadership: A Discussion with Colonel Everett Spain

Control: Human-generated

Podcast Description

Join us for another episode of *Inside West Point: Ideas That Impact*. In this episode, we speak with Colonel Everett Spain about leadership studies and inspiring leaders' impact the on talent development. **We discuss Spain's research on leader evaluations and retention patterns**, highlighting the importance of creating a positive culture. **We also touch on the role of battalion commanders in shaping future leaders** and the significance of innovation and intellectual development. Additionally, we discuss the importance of character and the ongoing research in this field.



Treatment: GPT-generated

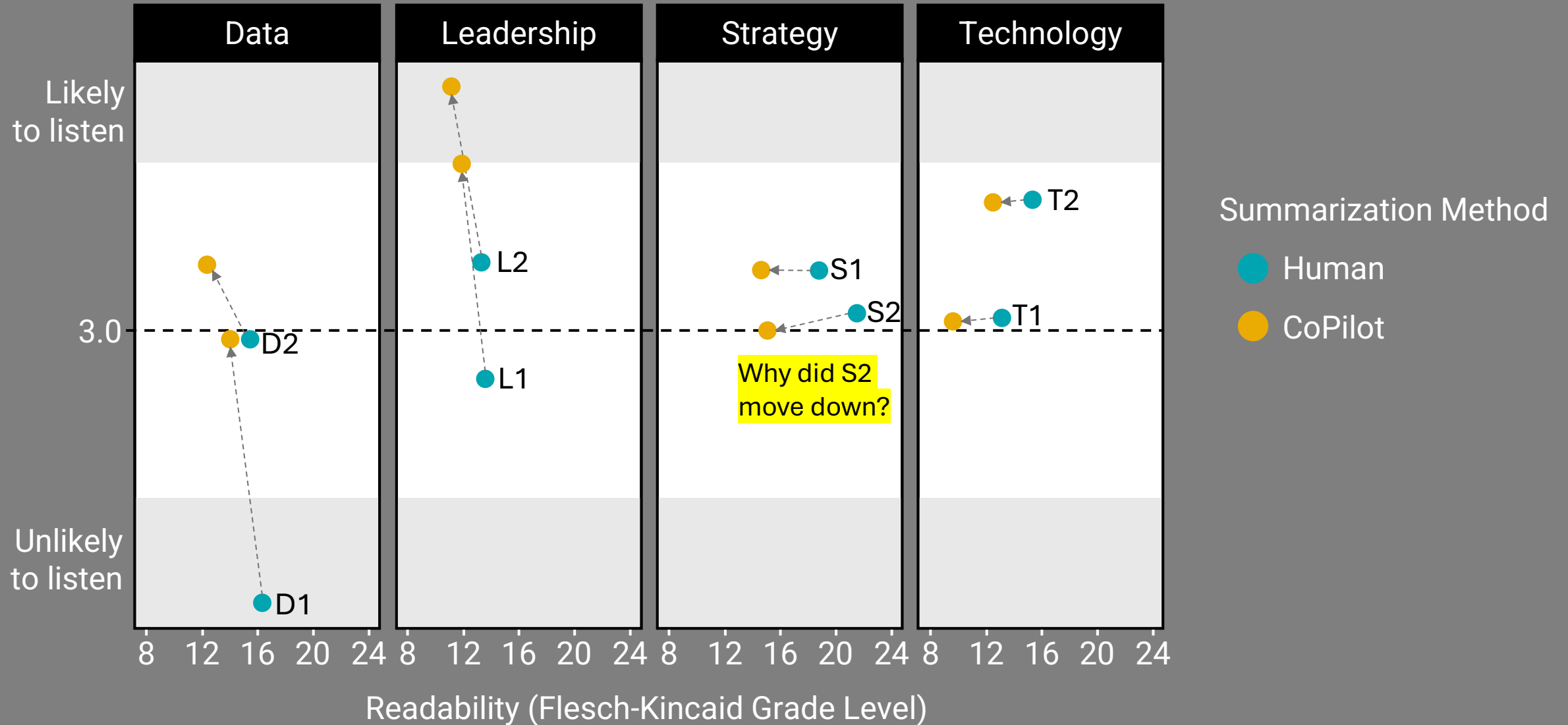
Battalion commanders shape more than readiness—they shape futures. Colonel Everett **Spain's research proves their influence on retention is decisive**, especially for high-potential officers. Leaders who fail to inspire risk losing talent when the Army needs it most. Command isn't just authority—it's a role-model crucible that determines whether tomorrow's leaders stay or walk.

Three Key Insights:

- Look for warning signs: **uninspiring leaders drain energy** and drive talent out.
- Strive for visible care and high standards—"you gotta care that much" about soldiers and families.
- Gain a sustainable advantage by **inspiring lieutenants to raise their hand for command** and stay in uniform.

Listen to Colonel Spain explain why inspiring leadership is the Army's most powerful weapon for retaining elite talent.

Intentions to Listen



S2: Building Cross-Cultural Competence Through Language and Cultural Immersion with Col. John Baskerville

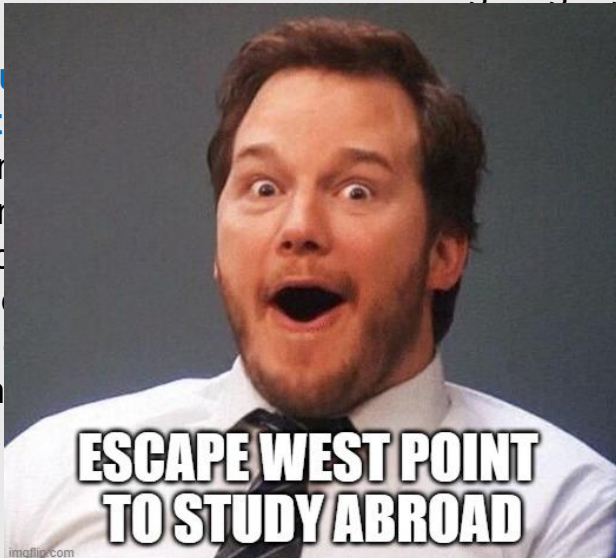
Control: Human-generated

Podcast Description

This episode of Inside West Point: Ideas that Impact, features an in-depth conversation with Colonel John Baskerville, head of the West Point Department of Foreign Languages, discussing the vital role of foreign language and cultural competence in military education and leader development. Baskerville highlights his extensive background in military service and foreign affairs, **emphasizing the importance of the semester abroad programs, international exchanges**, and the emerging challenges and innovations in teaching language.

The **disc** on cadet
on cadet
broader in
global con
experienc
contributi
advantag
operation

ese programs
ations, and the
ement in
transformative
immersion, their
the strategic
orary military



Treatment: GPT-generated

Coalition warfare demands more than firepower—it demands cultural fluency. Colonel John Baskerville warns: officers who ignore language and cross-cultural competence risk tactical missteps and strategic failure. From cadets navigating Persian taboos to leveraging German insights on recruiting, **leaders who embrace immersion sharpen judgment, build trust, and outthink adversaries in complex environments.**

Three Key Insights:

- **Look for cultural blind spots** that fracture alliances and erode mission legitimacy.
- Strive for deliberate immersion—**gain “another lens”** through language, history, and regional context.
- Gain operational advantage by **forging trust and interoperability** across diverse coalitions.

Listen to Colonel Baskerville explain why cultural competence is now a warfighting imperative for Army leaders.

People liked the CoPilot summaries, but...

CAUTION**Hallucinations
at Work**

Podcast Summarizer

50-Word Version (~50 words)

Afghanistan's withdrawal exposed a silent threat: leaders drowning in messy, unstructured data. Colonel Nicholas Clark warns that without "data-informed" judgment, commanders risk paralysis at the point of decision. Officers who master data literacy—asking the right questions, shaping raw inputs, and communicating insights—will weaponize information for speed, clarity, and mission success.

Three Key Insights:

- Look for chaotic **"data deserts"** and overwhelming raw feeds that obscure operational decisions.
- Strive for disciplined data literacy—ask sharper questions, format inputs, and enforce analytics in briefs.
- Gain faster decisions and readiness by turning sensors into solutions at the edge of combat.

Bottomline

Listen to Colonel Clark explain why data literacy is now a warfighting imperative for every Army leader.



COL Clark never said the phrase "data deserts" during his podcast

Limitations

- **Only surveyed cadets** → results may not generalize to civilians or other age groups
- **Podcasts differ from real briefings** → performance may not transfer to command updates
- **Human summaries weren't *optimized*** → Other writers might outperform the AI-based summary
- **GPT prompts weren't optimized** → better instructions or RAG could improve LLM outputs

Future Work

- Implement a RAG model to address traceability limitation
- Explore using click-rates instead of intentions to listen
- Examine actual meetings instead of just podcasts
- Experiment with different summarization prompts
- Compare AUTOCALIBRATE ratings with human ratings

Conclusion

- AI-generated podcast summaries can influence perceived relevance and intentions to listen, but their effectiveness depends on the topic.
- Due to hallucinations, humans must always verify AI-generated outputs. Retrieval Augmented Generation (RAG) should be explored.
- Future research should test effects across more realistic settings and continue refine how we use and evaluate AI, especially through prompt engineering.

Questions?