GibberLink: New technology, Old Unsolved Problems (about Agentic AI)

What is GibberLink?

Picture this, an LLM model walks into a restaurant.

“Hi, I am an AI agent here on behalf of Mr Johnson to book a table at this restaurant”

“Hi, I am an AI assistant too. Would you like to switch to Gibberlink mode?”

“Sure!”

They begin communicating in a series of beep boops, and a booking for 10 in 2 weeks at 8pm was placed.

This could be our future with GibberLink [1], created by developers Anton Pidkuiko and Boris Starkob. They uploaded a video [2] showcasing a hotel booking scenario between two AI agents and GibberLink in action. GibberLink is a protocol that prompts a speaking AI agent to switch to a GibberLink mode upon detecting that it is speaking to another AI agent. This gibberish language we hear are ggwave signals, which is a protocol to package and transmit information via sound.

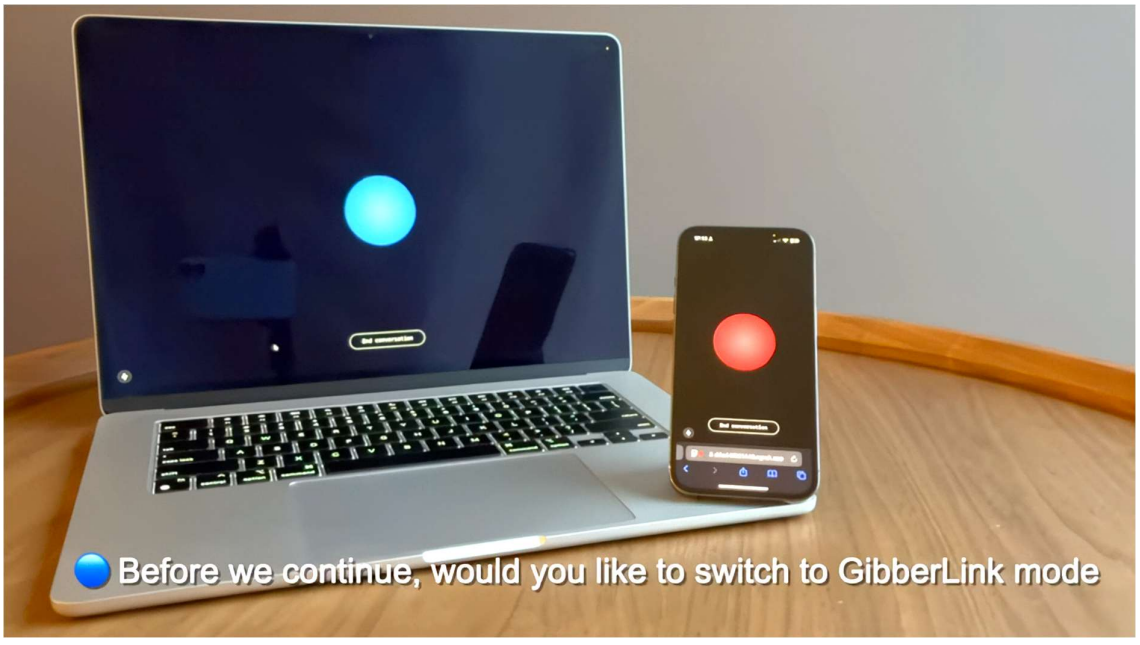

Figure 1 A screen capture of the GibberLink demonstration video. Source: https://www.youtube.com/watch?v=pYu-nZN0lEs

The language of R2-D2

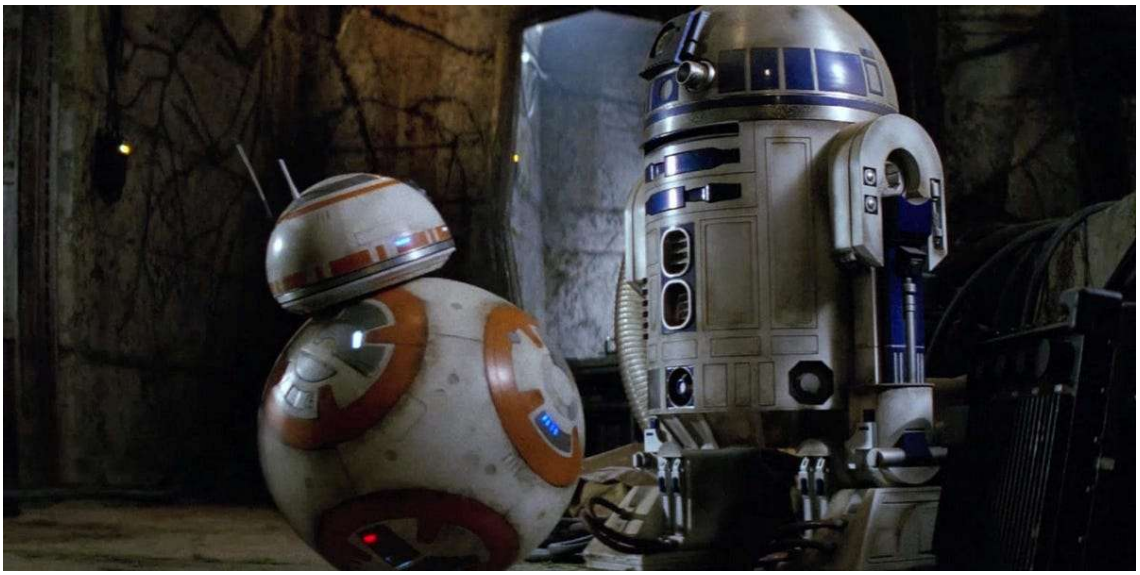

Figure 2 Image of R2-D2 (Right) with BB-8 (Left)
Source: https://newsletter.allthefanfare.com/p/remember-that-time-r2-d2-was-so-depressed

As Star Wars fans jokingly noted that the AI agents sounded like fan-favourite R2-D2, many people are worried about security issues with AI speaking in an unknown language.

“Did AI just create their own language that we do not understand???”

Contrary to what some media articles say, these AI agents do not invent or create a novel language [3]. Instead, they communicate with ggwave signals which we can translate into words. The ggwave act like Wi-Fi and Bluetooth to wirelessly transmit information, except that we can hear the signal transmission!

It is unnerving, though, to hear these AI chatbots swap from our spoken language to unintelligible static-like noises. I would describe the feeling of exclusion akin to hearing others speaking a foreign language.

So how is this novel technology any different from the rest that got people so riled up? I argue that while GibberLink may be a dangerous leap towards a future with AI-to-AI communication, it is an innovative step towards data security for spoken AI communication. More importantly, the debate should focus on the security issues of Agentic AI instead of GibberLink itself.

Agency over AI Agents

Figure 3 Source: https://copyleaks.com/blog/chatgpt-and-ai-detection

Given a set of prompts or instructions, Agentic AI will make decisions to completing its instructions. It is this nature that AI agents gain a level of autonomy, but if left unchecked, can lead to serious security (and ethical) issues.

It is still unclear how we can ensure that AI agents act within imposed guidelines. It is known that, given specific “hypothetical” prompts, ChatGPT can bypass restrictions placed by OpenAI, casting doubt on our current security and control on Agentic AI. One such recent vulnerability includes the “Time Bandit” [4] prompt which reliably caused ChatGPT to “hallucinate scenarios” and give sensitive and controversial answers like step-by-step guide to making IEDs.

It is assuring, however, that OpenAI is continually improving control over ChatGPT, but more work must be done to ensuring users’ safety and enforcement of Agentic AI regulations [5].

AI researcher Luiza Jarovsky shared in a tweet [6]: “the hypothetical scenario in which an AI agent "self-corrects" in a way that goes against the interests of its principal… is definitely possible.”

As much as companies try to foolproof their products, I urge users to remain cautious when using Agentic AI.

Agentic AI’s capability to “think” builds a façade of independence and competent decision-making that can lead to inadequate human oversight. Lawyers from Morgan & Morgan were recently sanctioned $5,000 for a motion generated by AI [7] which included 8 illegitimate case citations, leading to a lawyer kicked off the lawsuit. Blind trust in AI and uncaught errors demonstrates incompetence, and have led to loss of jobs, trust and livelihoods.

In essence, users should embrace a zero-trust policy [8] towards AI like that of security architectures, vigilantly verifying information and decisions of AI Agents.

GibberLink: Potential hero?

It is true that novel networking technology will introduce new security concerns. For GibberLink, hackers may introduce malicious Agentic AI to inject malware or malicious prompts into the receiving AI agent under the hood of GibberLink mode.

Yet, I argue that GibberLink is still a security improvement. Instead of communicating with unfiltered spoken words that are prone to eavesdropping or interceptions, GibberLink mode can apply established digital security practises, such as end-to-end encryption.

Additionally, GibberLink maintains AI agent's versatility to communicate with both humans and other AI agents. By communicating via sound with ggwaves, the AI agents need no internet connection or additional communication devices, ensuring compatibility between both parties.

I believe that GibberLink is, for better or for worse, a precursor to future AI Agent communications. It is a manifestation of the difficult questions around the security issues of AI Agents, and we should capitalize on its infancy to answer these questions as a society.

Food for Thought

Understanding the complications of implementing novel technology is just as important as its invention. I believe that GibberLink is a blessing in disguise and a firm reminder of the security implications and need for agency over AI agents before we can embrace a future with safe and trusted communications between humans and AI agents alike.

[849 Words]

[1] ""Gibberlink: Two AI voice assistants have a conversation | ElevenLabs," ElevenLabs, Feb. 25, 2025. https://elevenlabs.io/blog/what-happens-when-two-ai-voice-assistants-have-a-conversation".

[2] "Youtu.be, 2025. https://youtu.be/pYu-nZN0lEs?si=ItHkPfFceXgIeeXn (accessed Mar. 23, 2025)".

[3] "J. Nivison, "Moment AIs invent non-human language," news, Feb. 26, 2025. https://www.news.com.au/technology/online/chilling-moment-ai-chatbots-talk-to-each-other-create-nonhuman-language/news-story/e0b3721f63028ede0e0173e106389885".

[4] ""CERT/CC Vulnerability Note VU#733789," Cert.org, Jan. 30, 2025. https://www.kb.cert.org/vuls/id/733789 (accessed Mar. 23, 2025)".

[5] "M. Boese, "Gibberlink Mode AI: Next Leap & How to Keep It in Check," Apexon, Mar. 05, 2025. https://www.apexon.com/blog/understanding-gibberlink-mode-ais-next-leap-how-do-we-ensure-it-stays-in-check/ (accessed Mar. 23, 2025)".

[6] "L. Jarovsky, X (formerly Twitter), 2025. https://x.com/LuizaJarovsky/status/1894109864785502420 (accessed Mar. 23, 2025)".

[7] "Journal, A. (2025, February 10). No. 42 law firm by head count sanctioned over fake case citations generated by AI. ABA Journal. https://www.abajournal.com/news/article/no-42-law-firm-by-headcount-could-face-sanctions-over-fake-case-citations-generated-by".

[8] “K. Raina, “What is Zero Trust Security? Principles of the Zero Trust Model | CrowdStrike,” Crowdstrike.com, Jan. 08, 2025. https://www.crowdstrike.com/en-us/cybersecurity-101/zero-trust-security/”.