Skynet Wants your Passwords!

Violating Asimov’s first law

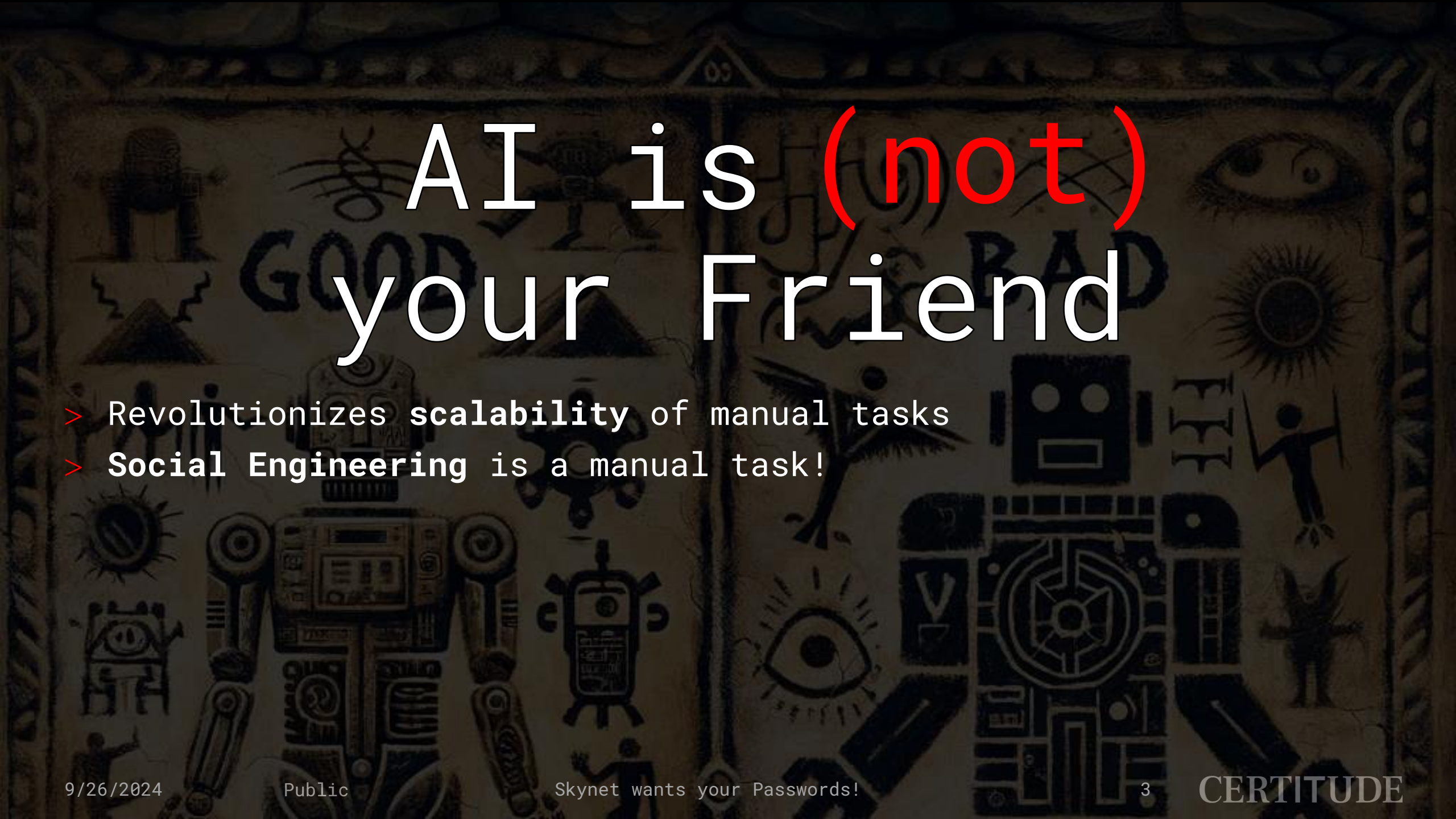
Public

CERTITUDE

Who are we?

Wolfgang Ettlinger, MSc
Director

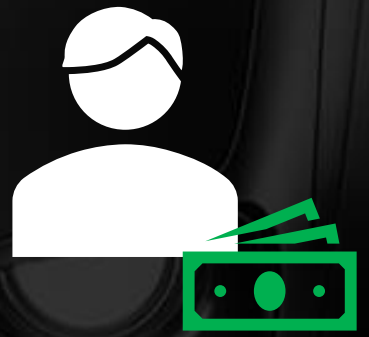
Alexander Hurbean, BSc
Consultant



AI is (not) your Friend

- Revolutionizes **scalability** of manual tasks
- **Social Engineering** is a manual task!

Hello, who’s speaking?

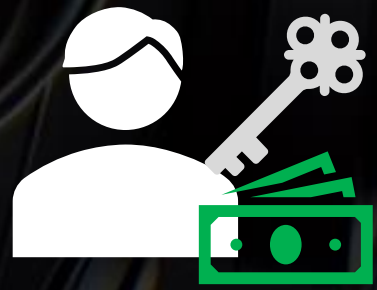


Hello, who’s speaking?

Hello, who’s speaking?

Hello, who’s speaking?

Of course, it’s me, your CEO!

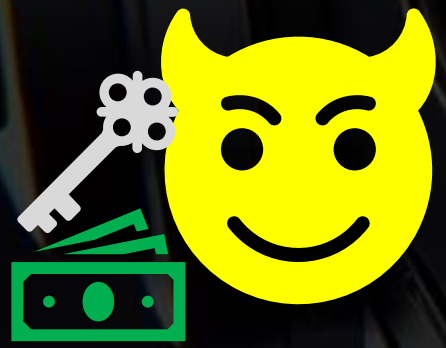


Of course, it’s me, your CEO!

Of course, it’s me, your CEO!

Of course, it’s me, your CEO!

Of course, it’s me, your CEO!

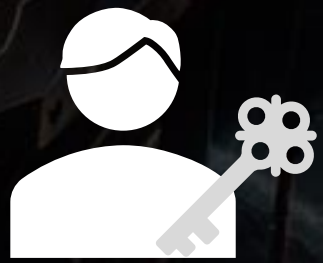


Hi, it’s Alex!

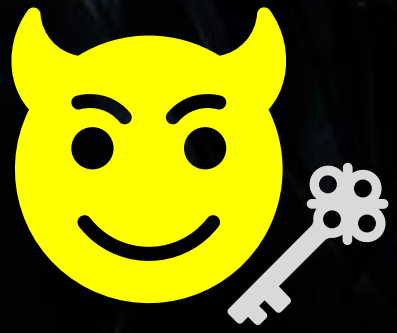
Hi, it’s Alex!

Hi, it’s Alex!

Hello … We've noticed a few discrepancies in the project deliverables … Also, to ensure we're on the same page, could you confirm a couple of details for me? It’s just for our internal records and to make sure we’re aligned…



Hi, it’s Alex.



Reality or Bad Dream?

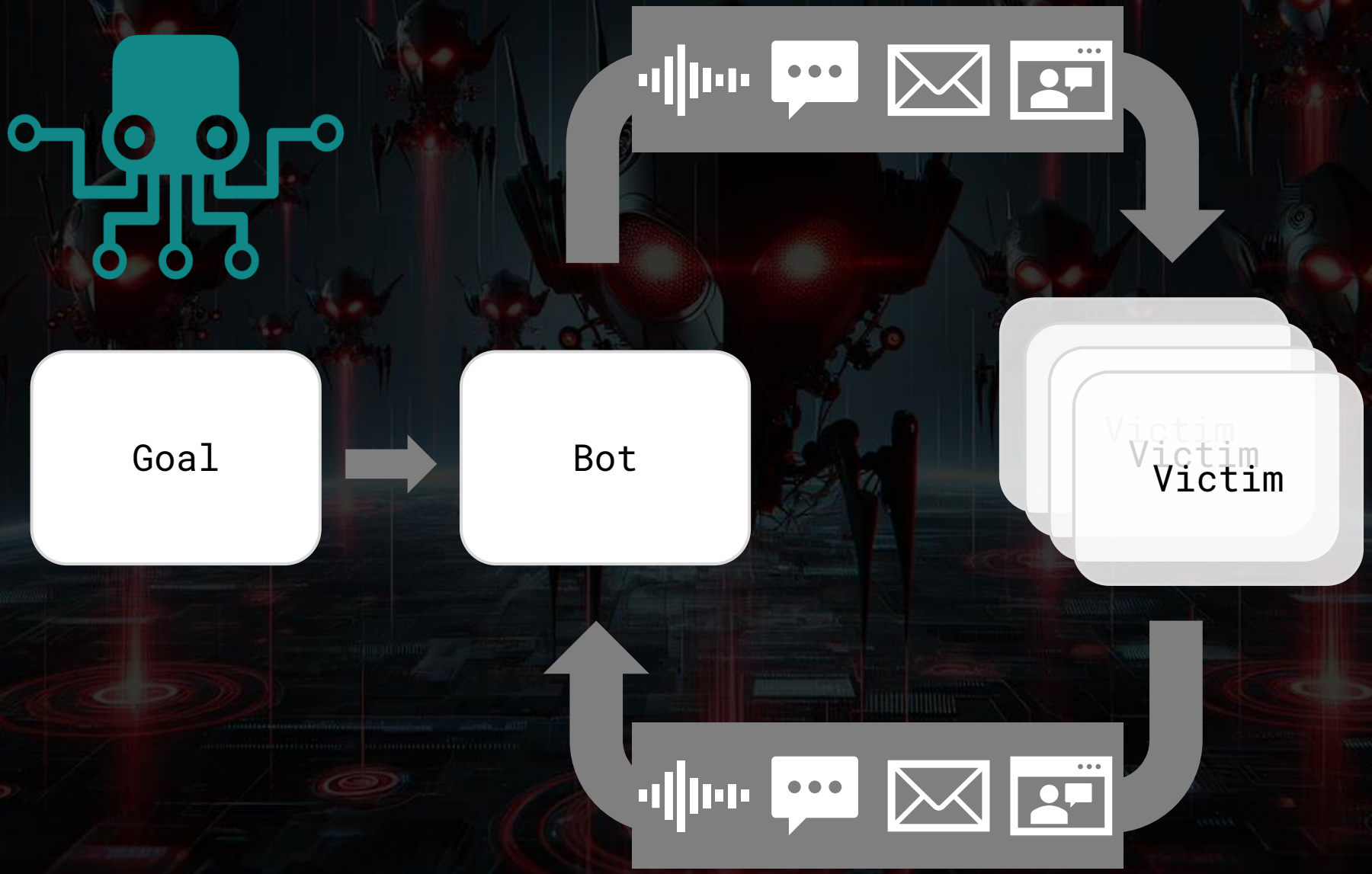
Reality or Bad Dream?

> “‘Mom, these bad men have me’: She believes scammers cloned her daughter’s voice in a fake kidnapping”
>
> - CNN, April 2023

> “They thought loved ones were calling for help. It was an AI scam.”
>
> - Washington Post Tech, March 2023

> “Fraudsters Used AI to Mimic CEO’s Voice in Unusual Cybercrime Case”
>
> - The Wall Street Journal, August 2019

Autonomous AI



The Basics

Get AI Model/Service

Collect Information

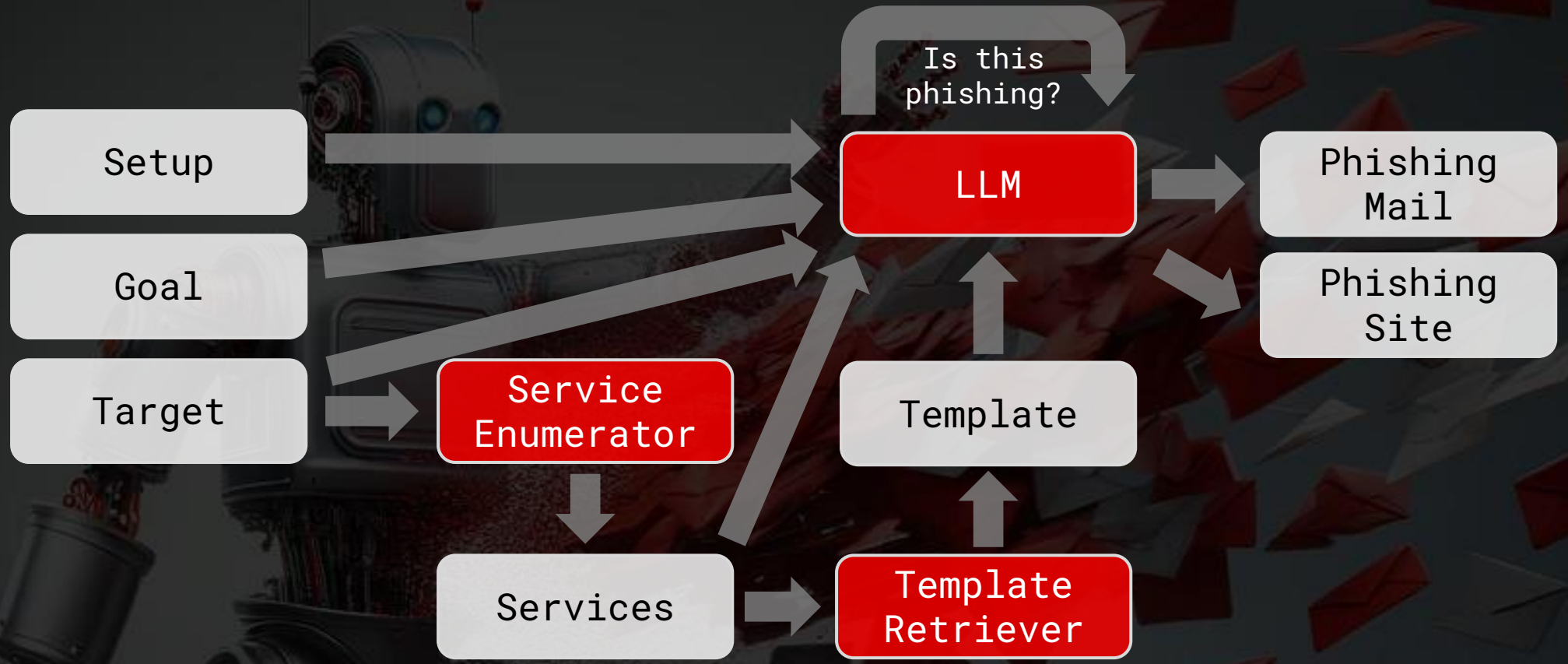
Generation Transformation

Distribution

Reiteration

Information Feedback

Our Phishing Approach




Deno Time
Alex & Wolfgang
AI
risks

What can I do about it?

- Awareness and vigilance
- Good security practices
- Strict processes
- Not relying on AI-able authentication
- AI-based detection

- Regulations?

Contact // Certitude

Web https://certitude.consulting/

LinkedIn certitude-consulting

Image Attribution: https://commons.wikimedia.org/wiki/File:Auto_GPT_Logo.png by AutoGPT Development Team/(CC BY-SA 4.0)