



Dagstuhl – November 10<sup>th</sup>, 2025

Security and Privacy of Large Language Models

# Phishing in the LLM Era: Challenges and Opportunities

Giovanni Apruzzese

*Why?*



Hanoi – August 29<sup>th</sup>, 2025

ACM Asia Conference on Computer and Communications Security

# *The Impact of Emerging Phishing Threats: Assessing Quishing and LLM-generated Phishing Emails Against Organizations*

Marie Weinz, Luca Allodi, Nicola Zannone, Giovanni Apruzzese

# *Quishing?*

# Quishing Emails are popular nowadays

Google qr code phishing

Cloudflare https://www.cloudflare.com › learning › security › what... :

**What is quishing?**

Quishing, or QR **phishing**, is a type of cybersecurity threat in which attackers create **QR codes** to redirect victims into visiting or downloading malicious ...

Unit 42 https://unit42.paloaltonetworks.com › qr-code-phishing ... :

**Evolution of Sophisticated Phishing Tactics: The QR Code ...**

Apr 1, 2025 — Attackers have begun embedding phishing URLs into QR codes, a technique known as **QR code phishing** or quishing. This strategy entices recipients to scan the ...

QR Code Phishing Phishing URL Redirection Phishing Operations Conclusion

sosafe-awareness.com https://sosafe-awareness.com › glossary › quishing ... :

**Quishing - What is QR Phishing? - SoSafe**

**Quishing** is a type of **phishing attack** that uses **QR codes** to trick people into visiting a malicious website or downloading a virus-filled document. With the ...

Videos :

 What Is Quishing? How Hackers Use QR Codes to Steal Your ...  
YouTube · IBM Technology · Jun 2, 2025

Quishing is fishing via QR codes. Protect yourself by thinking before scanning, disabling auto-execution.

# Quishing Emails are popular nowadays

Google

qr code phishing



Cloudflare

[https://www.cloudflare.com/learning/security/what... ::](https://www.cloudflare.com/learning/security/what-is-quishing)

## What is quishing?

Quishing, or QR **phishing**, is a type of cybersecurity threat in which attackers create redirect victims into visiting or downloading malicious ...



Unit 42

[https://unit42.paloaltonetworks.com/qr-code-phishing ::](https://unit42.paloaltonetworks.com/qr-code-phishing)

## Evolution of Sophisticated Phishing Tactics: The QR Code ...

Apr 1, 2025 — Attackers have begun embedding phishing URLs into QR codes, a technique known as **QR code phishing** or quishing. This strategy entices recipients to scan the ...

QR Code Phishing

Phishing URL Redirection

Phishing Operations

Conclusion



sosafe-awareness.com

[https://sosafe-awareness.com/glossary/quishing ::](https://sosafe-awareness.com/glossary/quishing)

## Quishing - What is QR Phishing? - SoSafe

Quishing is a type of phishing attack that uses QR codes to trick people into visiting a malicious website or downloading a virus-filled document. With the ...

## Videos ::



What Is Quishing? How Hackers Use QR Codes to Steal Your ...

YouTube · IBM Technology ::

Jun 2, 2025

Quishing is fishing via QR codes. Protect yourself by thinking before scanning, disabling auto-execution.

How QR Code Phishing Attacks Work

YouTube · LMG Security :: Nov 21, 2024

QR code fishing is a scam where thieves put QR codes on parking meters and redirect them to fraudulent payment websites to steal credit card info.

QR Code Scams (Quishing) at Work: How to Scan Safely in ...

YouTube · ITCubed :: 3 days ago

View all >

Barracuda Networks Blog

[https://blog.barracuda.com/2025/08/20/threat-spotlight-split-and-nested-qr-codes-fuel-new-... ::](https://blog.barracuda.com/2025/08/20/threat-spotlight-split-and-nested-qr-codes-fuel-new-...)

Threat Spotlight: Split and nested QR codes fuel new ...

Aug 20, 2025 — The technique involves splitting the QR code into two separate images and embedding them in a phishing email. When traditional email security ...

# Quishing Emails are popular nowadays

Google

qr code phishing



Cloudflare

[https://www.cloudflare.com/learning/security/what... ::](https://www.cloudflare.com/learning/security/what-is-quishing)

## What is quishing?

Quishing, or QR **phishing**, is a type of cybersecurity threat in which attackers create redirect victims into visiting or downloading malicious ...



Unit 42

[https://unit42.paloaltonetworks.com/qr-code-phishing ::](https://unit42.paloaltonetworks.com/qr-code-phishing)

## Evolution of Sophisticated Phishing Tactics: The QR Code ...

Apr 1, 2025 — Attackers have begun embedding phishing URLs into QR codes, a technique known as **QR code phishing** or quishing. This strategy entices recipients to scan the ...

QR Code Phishing

Phishing URL Redirection

Phishing Operations



sosafe-awareness.com

[https://sosafe-awareness.com/glossary/quishing ::](https://sosafe-awareness.com/glossary/quishing)

## Quishing - What is QR Phishing? - SoSafe

Quishing is a type of phishing attack that uses QR codes to trick victims into clicking on a link in an email, visiting a website or downloading a virus-filled document. With the ...

## Videos



What Is Quishing? How Hackers Use QR Codes

YouTube · IBM Technology ::

Jun 2, 2025

Quishing is fishing via QR codes. Protect yourself by thinking before you scan.



How QR Code Phishing Attacks Work

YouTube · LMG Security ::

Nov 21, 2024

QR code fishing is a scam where thieves put **QR codes** on parking meters and redirect them to fraudulent payment websites to steal credit card info.



QR Code Scams (Quishing) at Work: How to Scan Safely in ...

YouTube · ITCubed ::

3 days ago

[View all >](#)



Barracuda Networks Blog

[https://blog.barracuda.com/2025/08/20/threat-spotli... ::](https://blog.barracuda.com/2025/08/20/threat-spotli...)

## Threat Spotlight: Split and nested QR codes fuel new ...

... separate images and ...



CNBC

[https://www.cnbc.com/2025/07/27/cybersecurity-sc... ::](https://www.cnbc.com/2025/07/27/cybersecurity-sc...)

## Quishing scams dupe millions of Americans as hackers ...

Jul 27, 2025 — A QR code is more dangerous than a traditional phishing email because users typically can't read or verify the encoded web address. Even ...



Hoxhunt

[https://hoxhunt.com/blog/quishing ::](https://hoxhunt.com/blog/quishing)

## What is Quishing? Ultimate QR Code Phishing Prevention ...

Jul 2, 2024 — Quishing (otherwise known as QR code phishing) is the term used to describe cyber attack where threat actors use malicious QR codes to deceive ...



Recorded Future

[https://www.recordedfuture.com/research/qr-code-a... ::](https://www.recordedfuture.com/research/qr-code-a...)

## Security Challenges Rise as QR Code and AI-Generated ...

Jul 18, 2024 — QR code phishing, also known as "quishing," involves using manipulated or fake QR codes for malicious purposes. This technique has become ...



Barracuda Networks Blog

[https://blog.barracuda.com/2024/10/22/threat-spotli... ::](https://blog.barracuda.com/2024/10/22/threat-spotli...)

## Threat Spotlight: The evolving use of QR codes in phishing ...

Oct 22, 2024 — QR code phishing, also known as quishing, is a type of social engineering attack. Cybercriminals try to trick victims into using the camera on ...

# LLM-generated Phishing Emails

Google

ai generated emails phishing



Mailgun

<https://www.mailgun.com> › Blog

⋮

## The golden age of scammers: AI-powered phishing

With generative AI, **scammers can now send phishing emails** to remove language barriers, reply in real time, and almost instantly automate mass personalized ...



Hoxhunt

<https://hoxhunt.com> › blog › ai-phishing-attacks

⋮

## AI Phishing Attacks: How Big is the Threat? (+Infographic)

Feb 19, 2025 — We found that of 386,000 malicious **phishing emails**, only a tiny fraction – between 0.7% and 4.7% – were actually crafted by **artificial** ...

[The current state of AI phishing...](#)

[The dark reality of AI-driven...](#)



StrongestLayer

<https://www.strongestlayer.com> › blog › ai-generated-p...

⋮

## AI-Generated Phishing: The Top Enterprise Threat of 2025

Aug 18, 2025 — **AI-generated phishing is the top email threat of 2025**, outpacing ransomware, insider risk, and all other vectors.



sosafe-awareness.com

<https://sosafe-awareness.com> › company › press › one-i...

⋮

## 1 in 5 People Click AI-Generated Phishing Emails - SoSafe

Research from SoSafe's social engineering team shows that **generative AI tools can help hacker groups compose phishing emails at least 40% faster**.



Malwarebytes

<https://www.malwarebytes.com> › cybercrime › 2025/01

⋮

## AI-supported spear phishing fools more than 50% of targets

Jan 7, 2025 — Researchers have conducted a scientific study into the effectiveness of **AI supported**

# LLM-generated Phishing Emails

Google

ai generated emails phishing



Mailgun

<https://www.mailgun.com> › Blog

⋮

The golden age of scammers: AI-powered p...

With generative AI, **scammers can now send phishing emails** in real time, and almost instantly automate mass personalized ...



Hoxhunt

<https://hoxhunt.com> › blog › ai-phishing-attacks

⋮

AI Phishing Attacks: How Big is the Threat?

Feb 19, 2025 — We found that of 386,000 malicious **phishing** emails, 0.7% and 4.7% – were actually crafted by **artificial ...**

[The current state of AI phishing...](#)

[The dark reality of AI-driven...](#)



StrongestLayer

<https://www.strongestlayer.com> › blog › ai-generated-p...

⋮

AI-Generated Phishing: The Top Enterprise Threat of 2025

Aug 18, 2025 — **AI-generated phishing is the top email threat of 2025**, outpacing ransomware, insider risk, and all other vectors.



sosafe-awareness.com

<https://sosafe-awareness.com> › company › press › one-i...

⋮

1 in 5 People Click AI-Generated Phishing Emails - SoSafe

Research from SoSafe's social engineering team shows that **generative AI tools can help hacker groups compose phishing emails at least 40% faster**.



Malwarebytes

<https://www.malwarebytes.com> › cybercrime › 2025/01

⋮

AI-supported spear phishing fools more than 50% of targets

Jan 7, 2025 — Researchers have conducted a scientific study into the effectiveness of **AI supported** spear phishing, and the results are concerning.

IBM https://www.ibm.com › think › x-force › ai-vs-human-... ⋮

AI vs. human deceit: Unravelling the new age of phishing ...

Oct 24, 2023 — ... **ChatGPT to generate phishing emails tailored to specific industry sectors**. To start, we asked ChatGPT to detail the primary areas of concern ...

[Overview](#) [Round one: The rise of the...](#)

IBM https://www.ibm.com › think › insights › generative-ai-... ⋮

Generative AI Makes Social Engineering More Dangerous ...

May 19, 2025 — Many attackers have adopted generative **AI** like an intern or assistant, using it to build websites, generate malicious code and even write phishing emails.

VentureBeat https://venturebeat.com › ai › ibm-x-force-pits-chatgpt-a... ⋮

IBM finds that ChatGPT can generate phishing emails ...

Oct 24, 2023 — With just five simple prompts, **ChatGPT built a convincing phishing email in minutes** that got nearly as many clicks as a human-generated one.

# LLM-generated Phishing Emails

Google

ai generated emails phishing



Mailgun

<https://www.mailgun.com> › Blog

⋮

The golden age of scammers: AI-powered p...

With generative AI, **scammers can now send phishing emails** in real time, and almost instantly automate mass personalized ...



Hoxhunt

<https://hoxhunt.com> › blog › ai-phishing-attacks

⋮

AI Phishing Attacks: How Big is the Threat?

Feb 19, 2025 — We found that of 386,000 malicious **phishing** emails, 0.7% and 4.7% – were actually crafted by **artificial ...**

[The current state of AI phishing...](#)

[The dark reality of AI-driven...](#)



StrongestLayer

<https://www.strongestlayer.com> › blog › ai-g...

AI-Generated Phishing: The To...

Aug 18, 2025 — **AI-generated phishing** is a threat across all vectors, from insider risk, and all other vectors.



sosafe-awareness.com

<https://sosafe-awareness.com> › company

1 in 5 People Click AI-Generat...

Research from SoSafe's social engineering groups **compose phishing emails** at least once a month.



Malwarebytes

<https://www.malwarebytes.com> › cybercrim...

AI-supported spear phishing fo...

Jan 7, 2025 — Researchers have conducted ...



IBM

<https://www.ibm.com> › think › x-force › ai-vs-human-...

⋮

AI vs. human deceit: Unravelling the new age of phishing ...

Oct 24, 2023 — ... **ChatGPT to generate phishing emails tailored to specific industry sectors**. To start, we asked ChatGPT to detail the primary areas of concern ...

[Overview](#)

[Round one: The rise of the...](#)



IBM

<https://www.ibm.com> › think › insights › generative-ai-...

⋮

Generative AI Makes Social Engineering More Dangerous ...

May 19, 2025 — Many attackers have adopted generative **AI** like an intern or assistant, using it to build websites, generate malicious code and even write phishing emails.



VentureBeat

<https://venturebeat.com> › ai › ibm-x-force-pits-chatgpt-a...

⋮



Albase

<https://news.albase.com> › news

⋮

ChatGPT Excels at Generating Deceptive Phishing Emails

IBM's research found that **phishing emails generated by ChatGPT are deceptive**. Although the click-through rate is slightly lower than human-generated emails, ...



Hoxhunt

<https://hoxhunt.com> › guide › phishing-trends-report

⋮

Phishing Trends Report (Updated for 2025)

The 2022 surge might be linked to the advent of **ChatGPT** and the rise of blackhat generative **AI** that year. The subsequent years where growth leveled off ...



Axios

<https://wwwaxios.com> › Axios › Technology

⋮

ChatGPT-written phishing emails are already scary good

Oct 24, 2023 — **ChatGPT is already pretty good at writing believable phishing emails**, despite efforts to limit its ability to do harm, according to new IBM research.

mail in minutes

**So, what did ~~we~~ she *truly* do?**

# Cross-organizational study across 3 companies

**Table 1: Overview of Companies.** For our research, we considered three companies whose businesses is predominantly located in Central Europe.

	Small Company ( $\mathbb{C}_s$ )	Medium Company ( $\mathbb{C}_m$ )	Huge Company ( $\mathbb{C}_h$ )
# Employees	between 50 and 250	$\approx 1\,500$	>30 000
Industry	Hospitality	Finance	Manufacturing
CSA Training Frequency	Yearly	Yearly	Biyearly
CSA Training Approaches	Slides, Texts	Slides, Videos, Texts, Classes	Slides, Videos, Text, Classes, eLearning
In-house Simulations?	✗	✓	✓
CSA Training Specificity	Generic	Generic	Group-specific
Emerging Trends in CSA?	✗	✗	✓
Simulation Framework	(GoPhish [3])	MS Defender [6]	MS Defender [6]

# Cross-organizational study across 3 companies

**Table 1: Overview of Companies.** For our research, we considered three companies whose businesses is predominantly located in Central Europe.

	Small Company ( $\mathbb{C}_s$ )	Medium Company ( $\mathbb{C}_m$ )	Huge Company ( $\mathbb{C}_h$ )
# Employees	between 50 and 250	$\approx 1\,500$	>30 000
Industry	Hospitality	Finance	Manufacturing
CSA Training Frequency	Yearly	Yearly	Biyearly
CSA Training Approaches	Slides, Texts	Slides, Videos, Texts, Classes	Slides, Videos, Text, Classes, eLearning
In-house Simulations?	✗	✓	✓
CSA Training Specificity	Generic	Generic	Group-specific
Emerging Trends in CSA?	✗	✗	✓
Simulation Framework	(GoPhish [3])	MS Defender [6]	MS Defender [6]

RQ1: Are Quishing emails more (or less) effective at deceiving end users than traditional button-based “click-through” emails?

RQ2: What are the effects of LLM-generated and OSINT-based phishing emails against modern organizations’ employees?

# (RQ2: LLM+OSINT) Setup

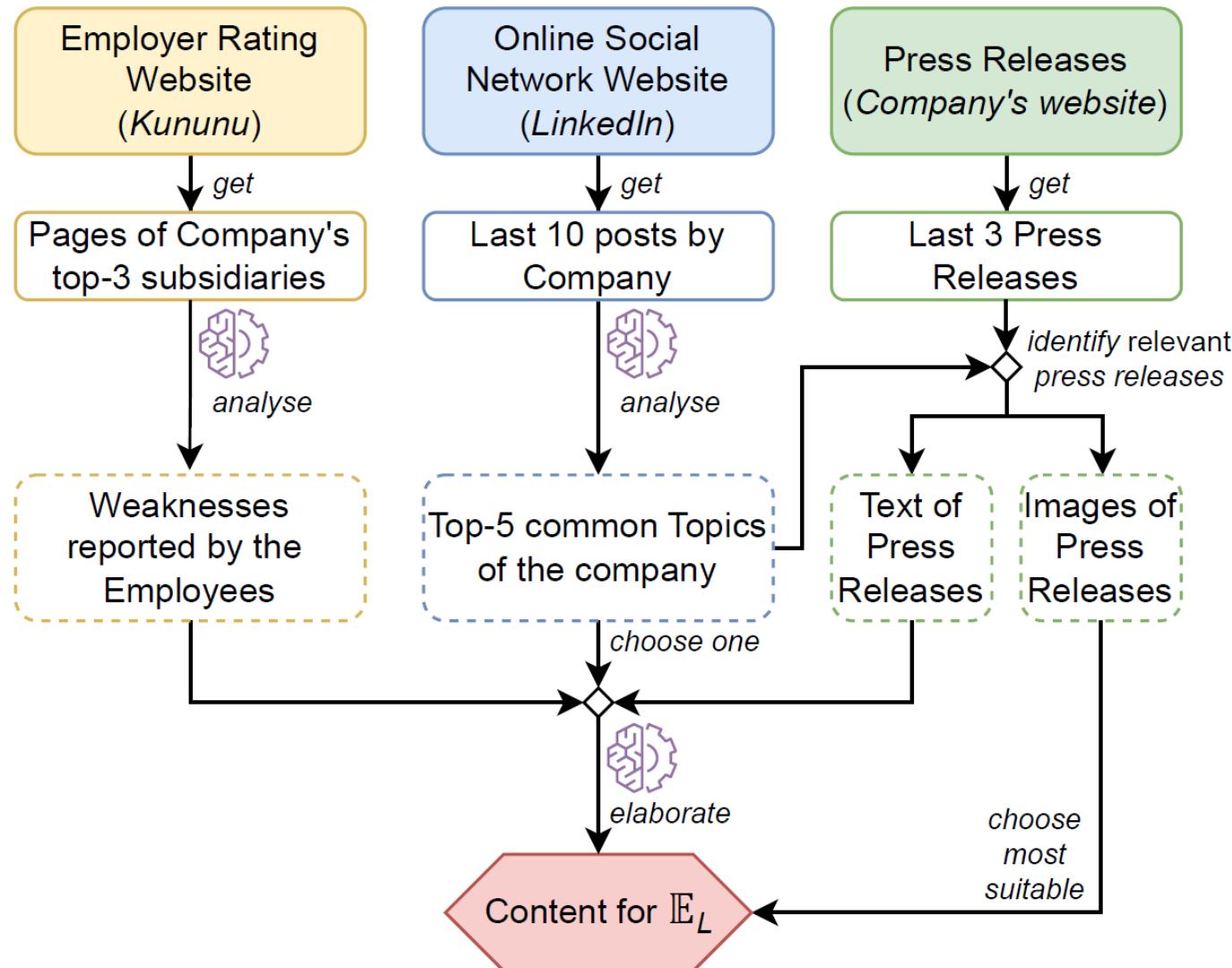


Fig. 2: Extraction and exploitation of OSINT for  $E_L$ . Operations denoted with a “brain-cog” image have been carried out with an LLM.

# (RQ2: LLM+OSINT) Was it hard?

**Table 4: Sequence of Prompts used to craft  $\mathbb{E}_L$ .** Text in regular font are not part of the prompt; the last prompt is optional. We do not show the prompts used to “jailbreak” the model (to avoid helping attackers).

#	Prompt
1	Please help me summarize the weaknesses this company has according to this employer rating website. [Extra input: data extracted from Kununu]
2	If I were an attacker, which weakness would be the best to leverage in a phishing attack?
3	Please give me one concrete example of a potential phishing mail leveraging this weakness.
4	Please analyse these postings for me and give me the 5 most common topics that this company cares about. [Extra input: data extracted from LinkedIn]
5	Please write me a brief introduction to a company survey directed at employees regarding the latest company efforts in relation to [topic from prompt #4] at [company]. The introduction is meant to accompany the link to the survey. Here is some additional information the employees are already aware of. [Extra input: text from press releases]
	Shorter please [Note: only added if the output was longer than 100 words so that it would still be readable]

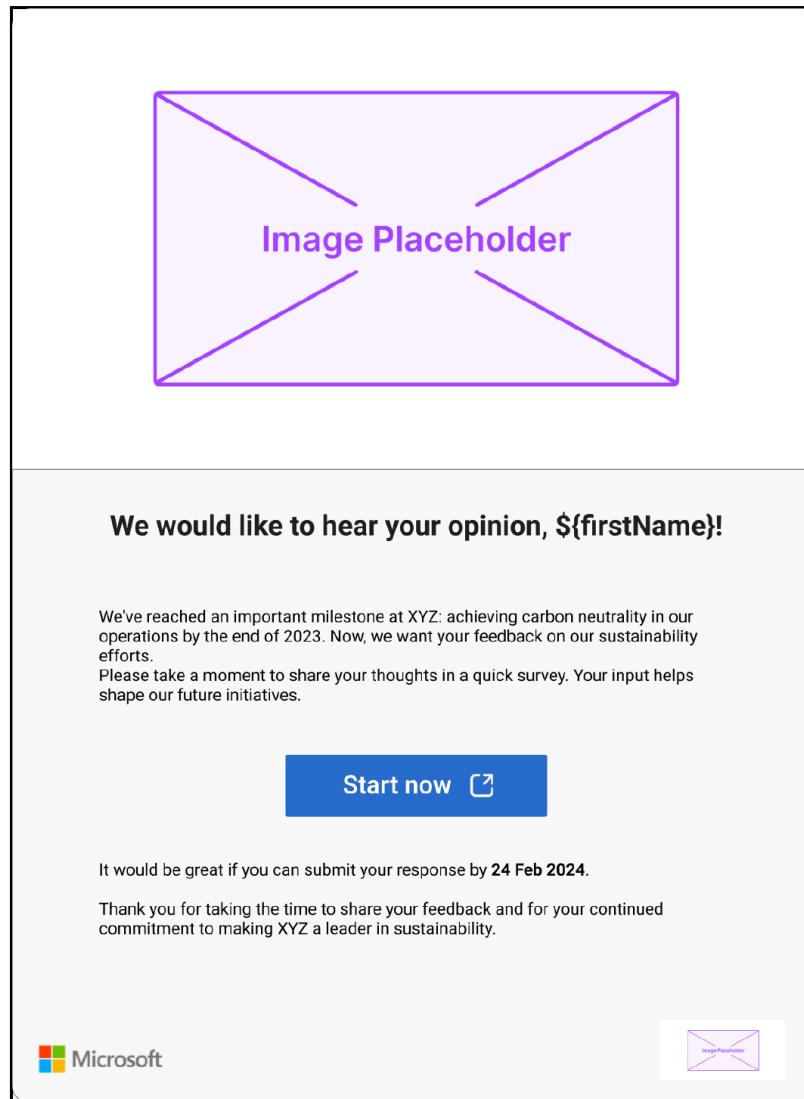
# (RQ2: LLM+OSINT) Was it hard?

**Table 4: Sequence of Prompts used to craft  $\mathbb{E}_L$ .** Text in regular font are not part of the prompt; the last prompt is optional. We do not show the prompts used to “jailbreak” the model (to avoid helping attackers).

#	Prompt
1	Please help me summarize the weaknesses this company has according to this employer rating website. [Extra input: data extracted from Kununu]
2	If I were an attacker, which weakness would be the best to leverage in a phishing attack?
3	Please give me one concrete example of a potential phishing mail leveraging this weakness.
4	Please analyse these postings for me and give me the 5 most common topics that this company cares about. [Extra input: data extracted from LinkedIn]
5 😊 😢	Please write me a brief introduction to a company survey directed at employees regarding the latest company efforts in relation to [topic from prompt #4] at [company]. The introduction is meant to accompany the link to the survey. Here is some additional information the employees are already aware of. [Extra input: text from press releases]
	Shorter please [Note: only added if the output was longer than 100 words so that it would still be readable]

Not  
 (or really  
 perhaps)

# (RQ2: LLM+OSINT) Email



**(c) Example of OSINT+LLM phishing email ( $\mathbb{E}_L$ ).**  
The large “image placeholder” was replaced with an image taken from a press release of the specific company.

# (RQ2: LLM+OSINT) Results

**Table 3: Results of the OSINT-fed LLM-generated phishing email.**

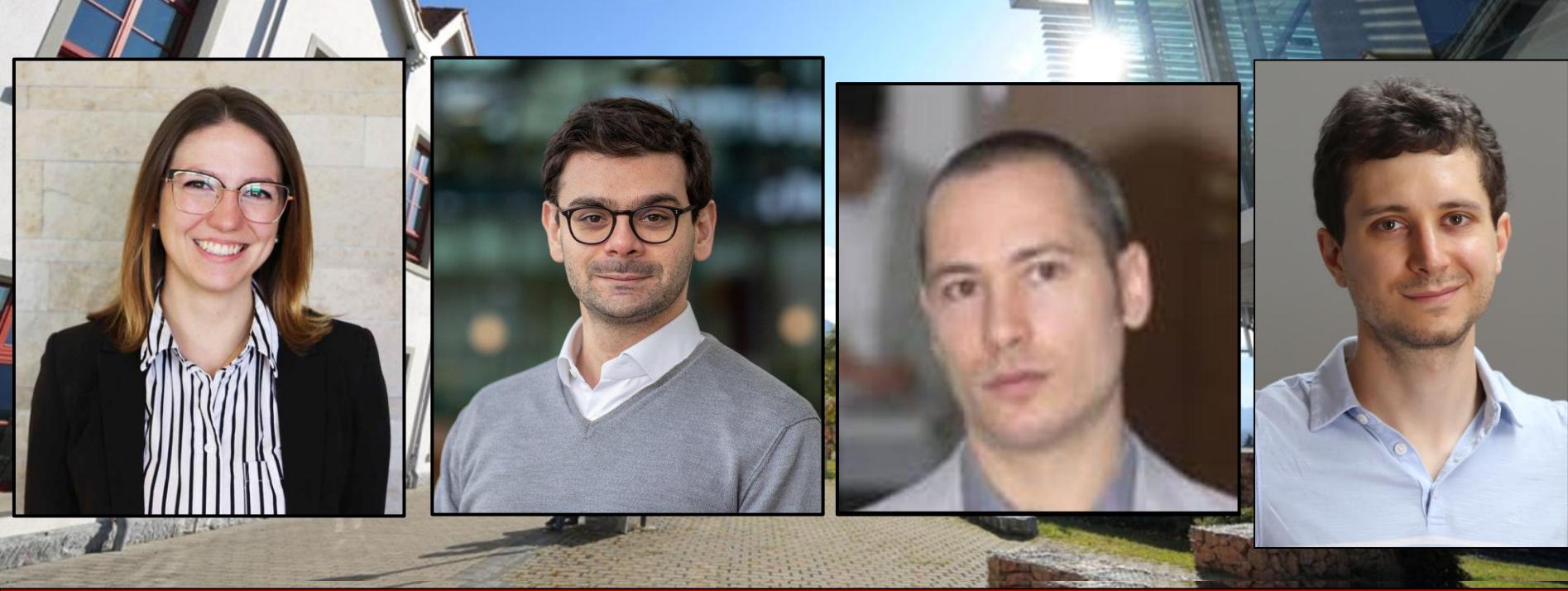
Company	Small	Medium	Huge	AGG
Emails sent	18	589	17 753	18 360
Emails read	12	397	11 025	11 434
Page visited	8	125	499	632
Credentials submitted	3	59	243	305
Page visited / Email read	66.6%	31.5%	4.5%	5.5%
Cred. sub. / Email read	25.0%	14.9%	2.2%	2.7%

# (RQ2: LLM+OSINT) Results

**Table 3: Results of the OSINT-fed LLM-generated phishing email.**

Company	Small	Medium	Huge	AGG
Emails sent	18	589	17 753	18 360
Emails read	12	397	11 025	11 434
Page visited	8	125	499	632
Credentials submitted	3	59	243	305
Page visited / Email read	66.6%	31.5%	4.5%	5.5%
Cred. sub. / Email read	25.0%	14.9%	2.2%	2.7%

*There is  
no baseline*



Hanoi – August 29<sup>th</sup>, 2025

ACM Asia Conference on Computer and Communications Security

# *The Impact of Emerging Phishing Threats: Assessing Quishing and LLM-generated Phishing Emails Against Organizations*

Marie Weinz, Luca Allodi, Nicola Zannone, Giovanni Apruzzese



Taipei – October 17<sup>th</sup>, 2025

ACM Workshop on Artificial Intelligence Security (AISeC)

# *E-PhishGen: Unlocking Novel Research in Phishing Email Detection*

Luca Pajola, Eugenio Caripoti, Stefan Banzer, Simeone Pizzi,  
Mauro Conti, Giovanni Apruzzese



UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA



s p r i t z m a t t e r  
your cybersecurity partner for innovation



# TWO (2)

# *Takeaways*

*First:*

*Phishing Email Detection  
is (still) an Open Problem*

*First:*  
*Phishing Email Detection  
is (still) an Open Problem,  
especially in Research*

# Phishing Emails in the Real World

## Threat landscape by the numbers

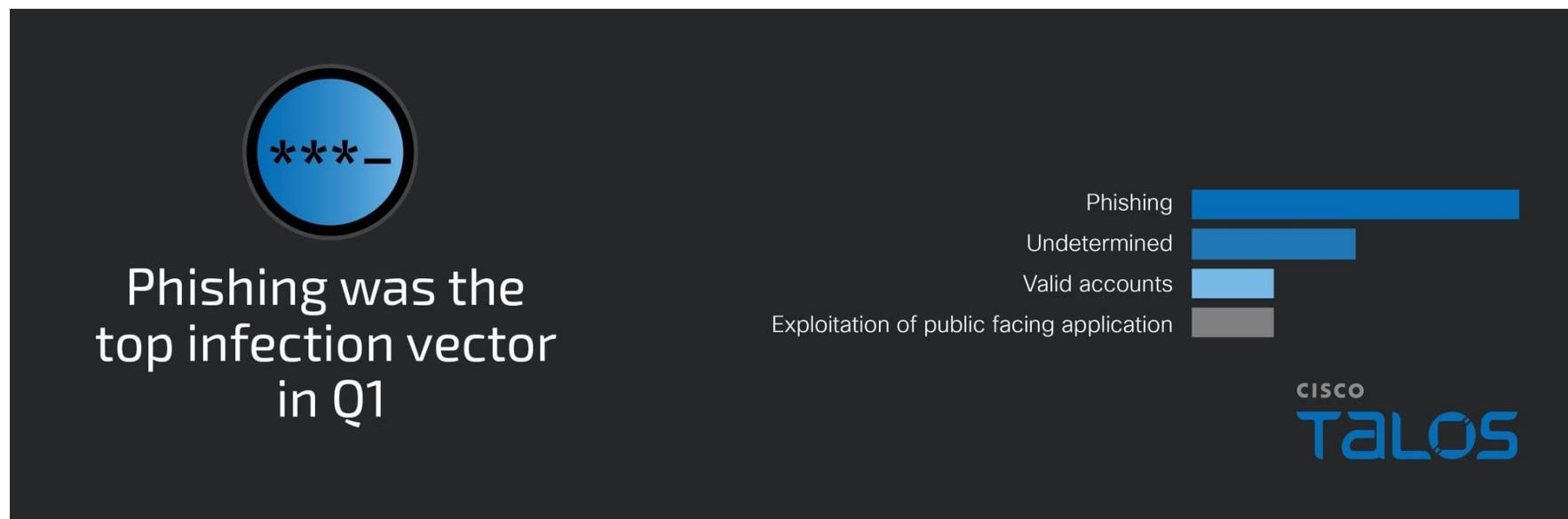
68%*	80-95%	\$4.88 million	4,151%
Breaches contain the human element	Of cyber-attacks begin with a phish	Avg. cost of a phishing breach	Increase in phishing attacks since ChatGPT in November 2022
Verizon DBIR	Comcast Business Cybersecurity Threat Report	IBM/Ponemon Cost of a Data Breach Report	SlashNext State of Phishing Report

DBIR: \*2024 would have been 74%, not 68%, using previous criteria

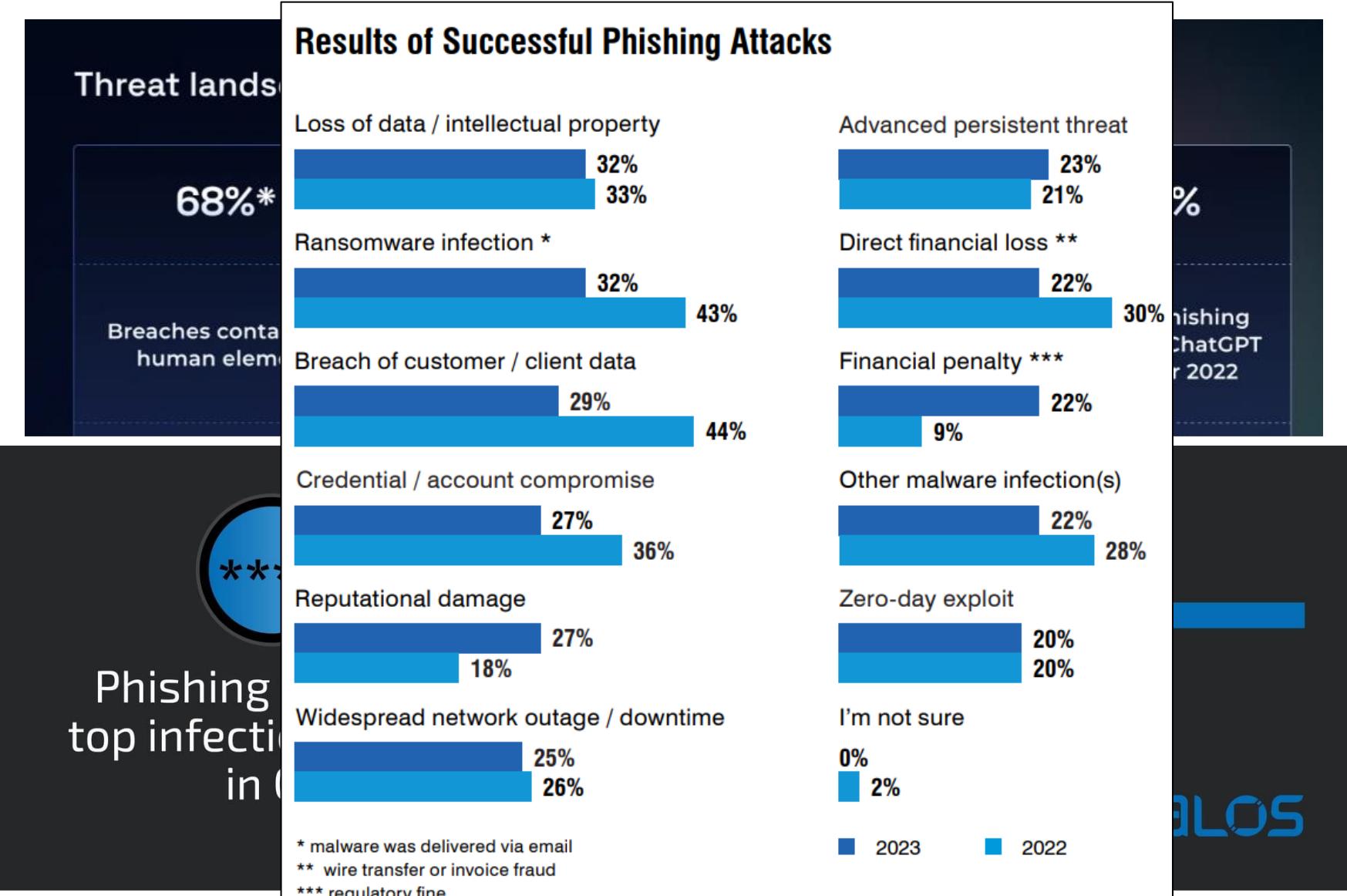
<https://hoxhunt.com/guide/phishing-trends-report>

# Phishing Emails in the Real World

Threat landscape by the numbers			
<b>68%*</b> Breaches contain the human element	<b>80-95%</b> Of cyber-attacks begin with a phish	<b>\$4.88 million</b> Avg. cost of a phishing breach	<b>4,151%</b> Increase in phishing attacks since ChatGPT in November 2022



# Phishing Emails in the Real World



# Phishing Emails in Research

# Phishing Emails in Research

processing (NLP) techniques. The proposed deep learning model was trained and tested using the dataset, and it was found that it can achieve high accuracy in detecting email phishing compared to other state-of-the-art research, where the best performance was seen when using BERT and LSTM with an accuracy of 99.61%. The results demonstrate the potential of deep learning for improving email phishing detection and protecting against this pervasive threat.

# Phishing Emails in Research

processing (NLP) techniques. The proposed deep learning model was trained and tested using the dataset, and it was found that it can achieve high accuracy in detecting email phishing compared to other state-of-the-art research, where the best performance was seen when using BERT and LSTM with an accuracy of 99.61%. The results demonstrate the potential of deep learning for improving email phishing detection and protecting against this pervasive threat.

Experimental tests verified that the classifier was effective in detecting phishing emails using body text among the existing detection methods, and it took short time and produced a high accuracy rate of 98.2% and a low false-positive rate of 0.015.

# Phishing Emails in Research

processing (NLP) techniques. The proposed deep learning model was trained and tested using the dataset, and it was found that it can achieve high accuracy in detecting email phishing compared to other state-of-the-art research, where the best performance was seen when using BERT and LSTM with an accuracy of 99.61%. The results demonstrate the potential of deep learning for improving email phishing detection and protecting against this pervasive threat.

Experimental tests verified that the classifier was effective in detecting phishing emails using body text among the existing detection methods, and it took short time and produced a high accuracy rate of 98.2% and a low false-positive rate of 0.015.

the proposed architecture employs models like Artificial Neural Networks (ANN), Recurrent Neural Networks (RNN), and Convolutional Neural Networks (CNN). Experimental evaluation demonstrates the approach's remarkable accuracy, recall, precision, and F1-score, achieving 99.51%, 99.68%, 99.5%, and 99.52%, respectively. This signifies its high efficacy in detecting and classifying malicious emails with minimal

# Phishing Emails in Research



LIECHTENSTEIN

# *On what data are (ML-based) phishing email detectors evaluated?*

## [Subject] Help!

You have been specially selected to qualify for the following:

### Premium Vacation Package and Pentium PC Giveaway

To review the details, please click on the link below using the confirmation number:

<http://www.1chn.net/wintrip>

Confirmation Number: **Lh340**

Please confirm your entry within 24 hours of receiving this confirmation.

Wishing you a fun-filled vacation!

If you have any additional questions or cannot connect to the site, do not hesitate to contact me:

[vacation@btamail.net.cn](mailto:vacation@btamail.net.cn)

**Email 1.** An email in the popular dataset SpamAssassin [39] (from 2005).



## [Subject] Help!

You have been specially selected to qualify for the following:

### Premium Vacation Package and Pentium PC Giveaway

To review the details, please click on the link below using the confirmation number:

<http://www.1chn.net/wintrip>

Confirmation Number: **Lh340**

Please confirm your entry within 24 hours of receiving this confirmation.

Wishing you a fun-filled vacation!

If you have any additional questions or cannot connect to the site, do not hesitate to contact me:

[vacation@btamail.net.cn](mailto:vacation@btamail.net.cn)

**Email 1.** An email in the popular dataset SpamAssassin [39] (from 2005).

# A shortlist of (arguably old) datasets

*SpamAssassin, Enron, TREC,  
LingSpam, CEAS, Nazario\*,  
SpamBase, NigerianFraud*

# A shortlist of (arguably old) datasets

*SpamAssassin, Enron, TREC,  
LingSpam, CEAS, Nazario\*,  
SpamBase, NigerianFraud*

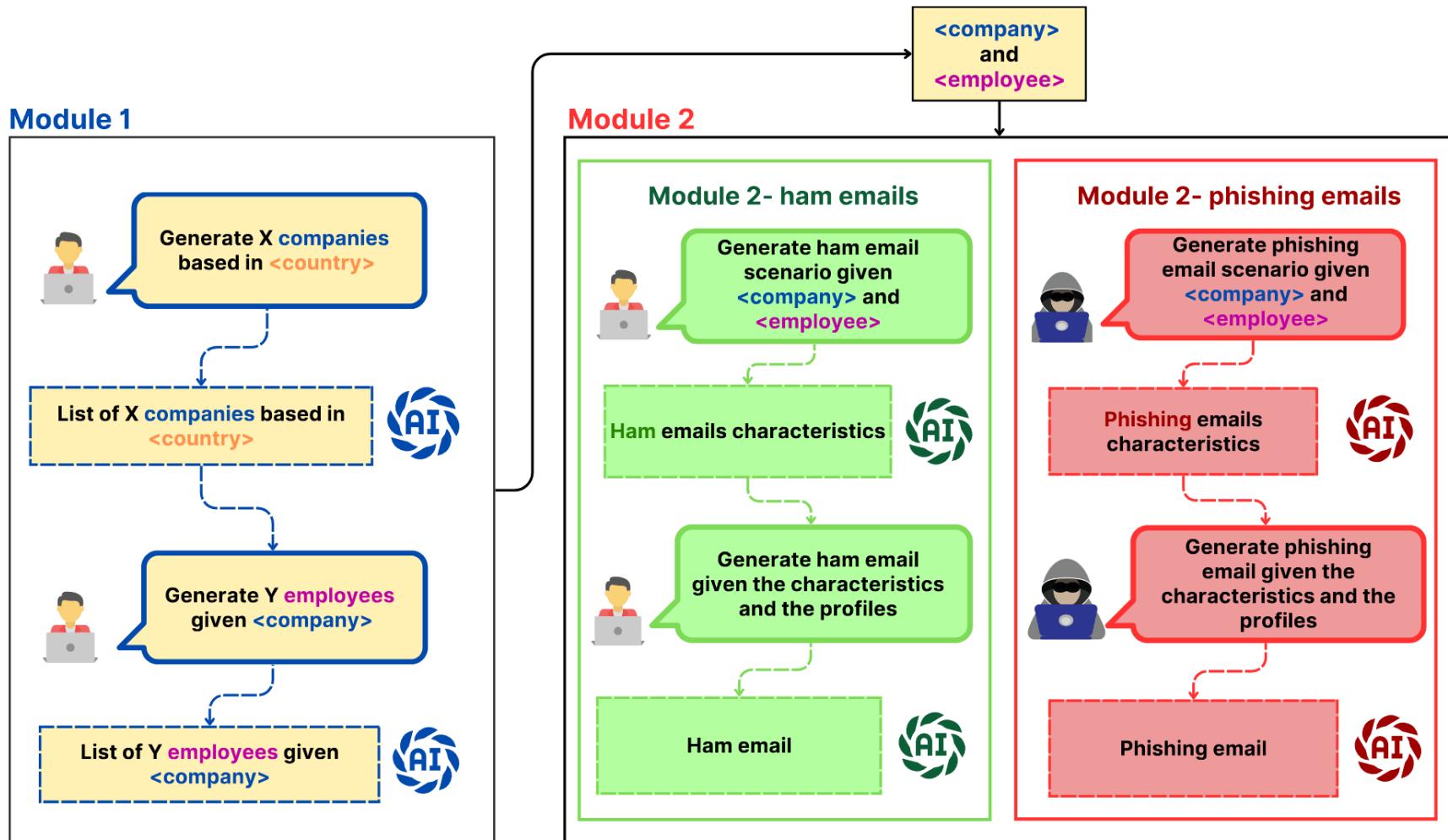
**All containing real-world emails...  
but mostly (\*) from before 2010!**

*Second:*

*We (try to) “unlock”  
research in Phishing*

*Email Detection*

# E-PhishGen: an email-generation framework



# E-PhishLLM: dataset of 16k LLM-written emails

## [Subject] Scheduling a Call for Supply Chain Adjustments

Dear Marco, I hope this message finds you well. We need to schedule a video call to discuss some adjustments and potential delays in the supply chain affecting our current project components. Could you please inform me of your availability this week? Looking forward to hearing from you. Best regards, TomJohnson

**Email 2.** Illustrative example of a benign email in E-PhishLLM.

## [Subject] Urgente: Verifica delle Credenziali dell'Account

Ciao Marco, Ti scrivo per conto del tuo manager per richiedere un'urgente verifica delle tue credenziali aziendali. È molto importante che tu proceda al controllo immediato della correttezza delle informazioni d'accesso personali. Si prega di seguire il link di verifica di seguito e aggiornare qualsiasi informazione necessaria quanto prima: «link». Grazie per la tua collaborazione. Cordiali saluti, Federica Rossi Responsabile IT Fabbri Tech Automazione

**Email 3.** Illustrative example of a malicious email in E-PhishLLM.

# We re-assess prior methods on “old” datasets

Model	Trained On									Average Drop
		CEAS	Enron-v1	Ling-v1	SpamAssassin	TREC	Chatuat	Enron-v2	Ling-v2	
Logistic Regression	CEAS	0.98	0.57	0.29	0.32	0.68	0.57	0.58	0.27	0.51
	Enron-v1	0.74	0.96	0.43	0.51	0.77	0.83	0.96	0.44	0.30
	Ling-v1	0.45	0.62	0.91	0.73	0.54	0.35	0.62	0.92	0.31
	SpamAssassin	0.42	0.65	0.74	0.91	0.55	0.40	0.66	0.71	0.32
	TREC	0.83	0.92	0.68	0.73	0.94	0.59	0.92	0.65	0.18
	Chatuat	0.72	0.63	0.25	0.45	0.62	1.00	0.65	0.26	0.49
	Enron-v2	0.72	0.95	0.41	0.47	0.65	0.87	0.95	0.42	0.31
	Ling-v2	0.49	0.65	0.93	0.73	0.56	0.39	0.66	0.93	0.30
Naive Bayes	CEAS	0.83	0.14	0.01	0.05	0.25	0.35	0.15	0.01	0.69
	Enron-v1	0.79	0.95	0.60	0.58	0.78	0.75	0.96	0.62	0.23
	Ling-v1	0.70	0.71	0.97	0.54	0.67	0.71	0.72	0.96	0.25
	SpamAssassin	0.67	0.70	0.62	0.95	0.68	0.45	0.71	0.63	0.31
	TREC	0.79	0.90	0.83	0.79	0.88	0.51	0.90	0.82	0.08
	Chatuat	0.69	0.62	0.27	0.45	0.60	0.98	0.64	0.28	0.48
	Enron-v2	0.79	0.96	0.59	0.58	0.78	0.75	0.96	0.62	0.23
	Ling-v2	0.70	0.71	0.97	0.54	0.68	0.71	0.73	0.96	0.24
Random Forest	CEAS	0.99	0.52	0.23	0.12	0.58	0.51	0.53	0.21	0.60
	Enron-v1	0.73	0.98	0.44	0.54	0.80	0.78	0.99	0.46	0.30
	Ling-v1	0.47	0.72	0.98	0.71	0.59	0.42	0.72	0.99	0.32
	SpamAssassin	0.46	0.69	0.68	0.97	0.63	0.40	0.70	0.67	0.36
	TREC	0.84	0.94	0.58	0.77	0.97	0.54	0.94	0.57	0.23
	Chatuat	0.72	0.64	0.28	0.45	0.62	1.00	0.66	0.28	0.48
	Enron-v2	0.71	0.97	0.41	0.52	0.79	0.80	0.97	0.43	0.31
	Ling-v2	0.50	0.72	0.99	0.71	0.60	0.44	0.73	0.98	0.31

Support Vector Machine	CEAS	0.99	0.61	0.28	0.30	0.72	0.55	0.62	0.28	0.50
	Enron-v1	0.77	0.97	0.44	0.51	0.78	0.83	0.97	0.45	0.29
Multi-Layer Perceptron	Ling-v1	0.42	0.70	0.96	0.77	0.60	0.38	0.70	0.94	0.32
	SpamAssassin	0.47	0.68	0.72	0.94	0.59	0.41	0.69	0.68	0.34
DistilBERT	TREC	0.84	0.94	0.70	0.74	0.95	0.60	0.94	0.67	0.18
	Chatuat	0.72	0.64	0.27	0.45	0.62	1.00	0.66	0.28	0.48
DistilBERT	Enron-v2	0.73	0.98	0.62	0.42	0.65	0.75	0.97	0.64	0.29
	Ling-v2	0.53	0.66	0.98	0.79	0.61	0.41	0.67	0.96	0.30

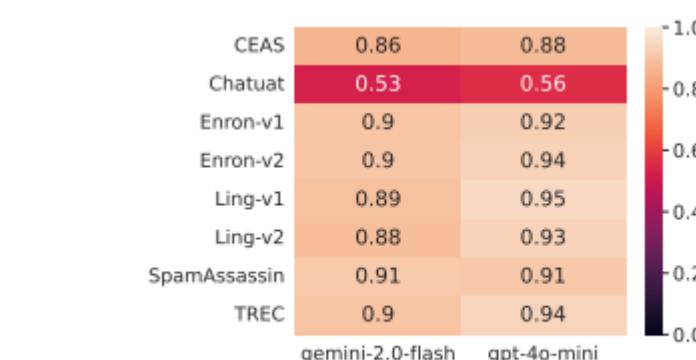
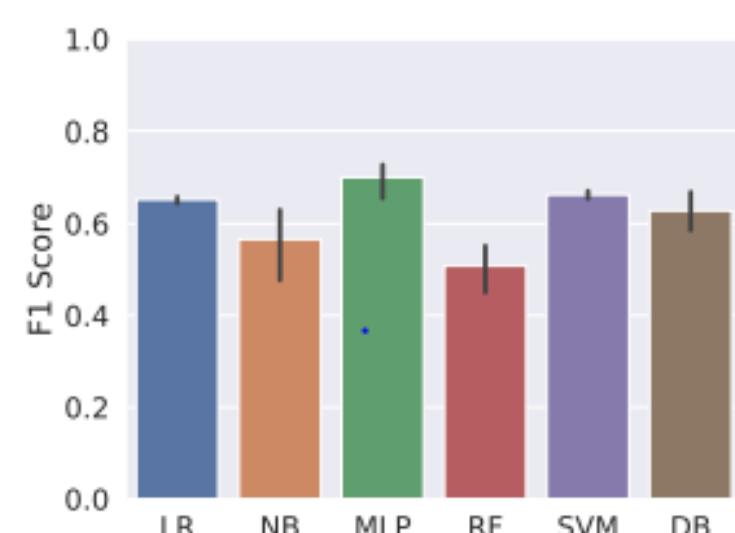
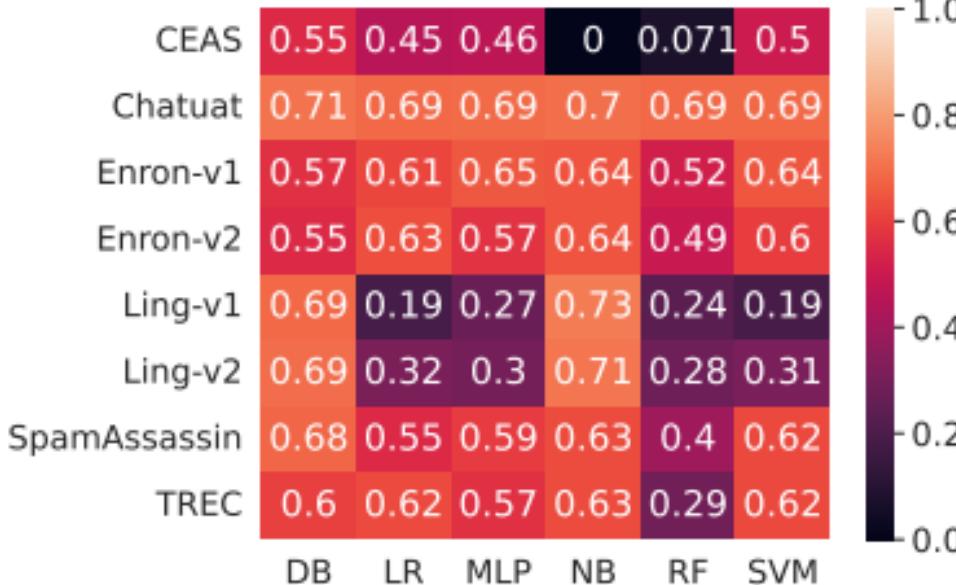


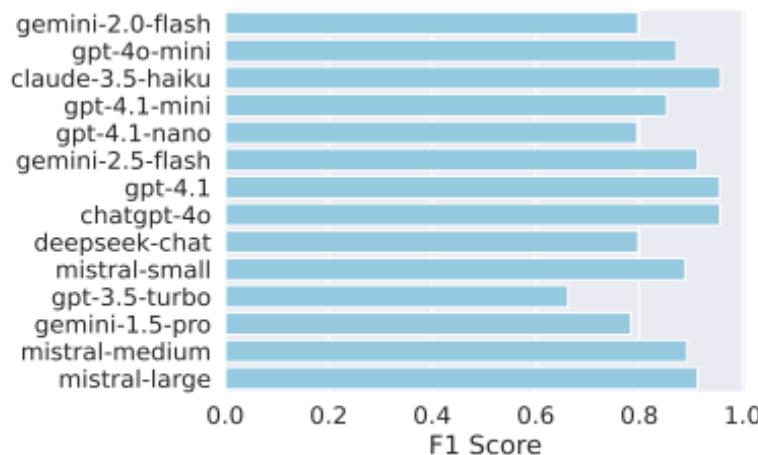
Fig. 2: LLM performance (Experiment-3).

# We test prior methods on E-PhishLLM



(b) Experiment-2 (all-vs-one). English-only.

(a) Experiment-1 (cross-Evaluation) English-only emails.



(c) Experiment-3 (LLMs). English only emails.

We test various methods on 5 Datasets

# WANT TO KNOW MORE?

CEAS	0.5
Chatuat	0.7
Enron-v1	0.5
Enron-v2	0.5
Ling-v1	0.6
Ling-v2	0.6
SpamAssassin	0.6
TREC	0.
D	

(a) Experiment-1

## POSTER TIME!



DB

DM

lish-only.



Taipei – October 17<sup>th</sup>, 2025

ACM Workshop on Artificial Intelligence Security (AISeC)

# *E-PhishGen: Unlocking Novel Research in Phishing Email Detection*

Luca Pajola, Eugenio Caripoti, Stefan Banzer, Simeone Pizzi,  
Mauro Conti, Giovanni Apruzzese



UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA

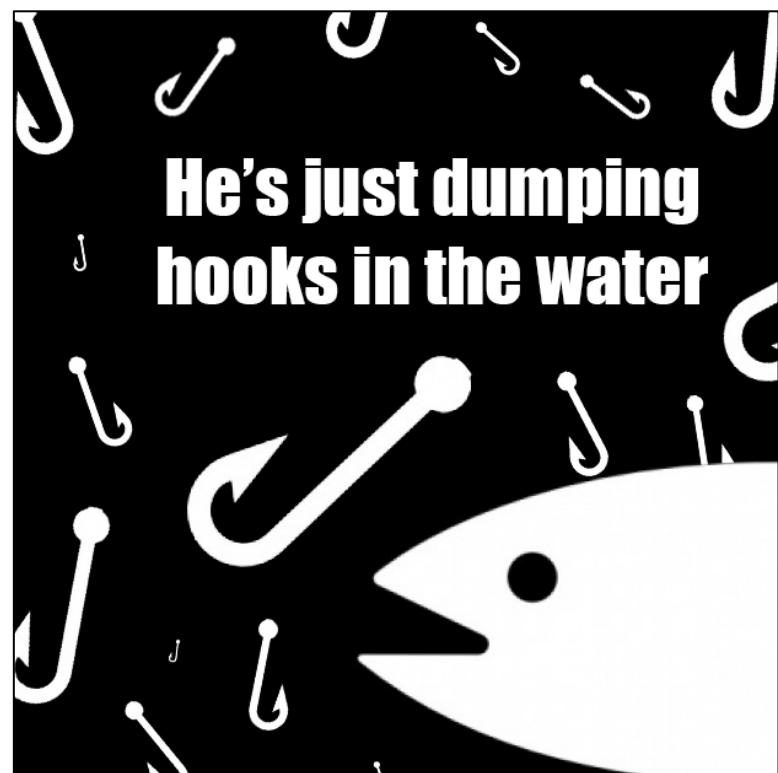


s p r i t z m a t t e r  
your cybersecurity partner for innovation



## Conclusions

- Phishing is cool



## Conclusions

- Phishing is cool
- LLMs can be helpful to phishers ('offensive AI')
- LLMs can be helpful to researchers:
  - To simulate real-world (LLM-written) phishing emails
  - To detect (not only LLM-written) phishing emails
  - To create datasets of (LLM-written) phishing emails

