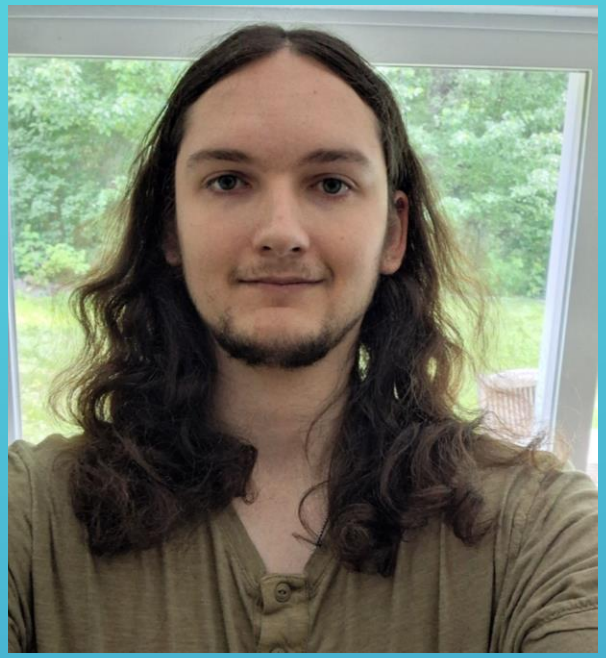
Phishing Detection

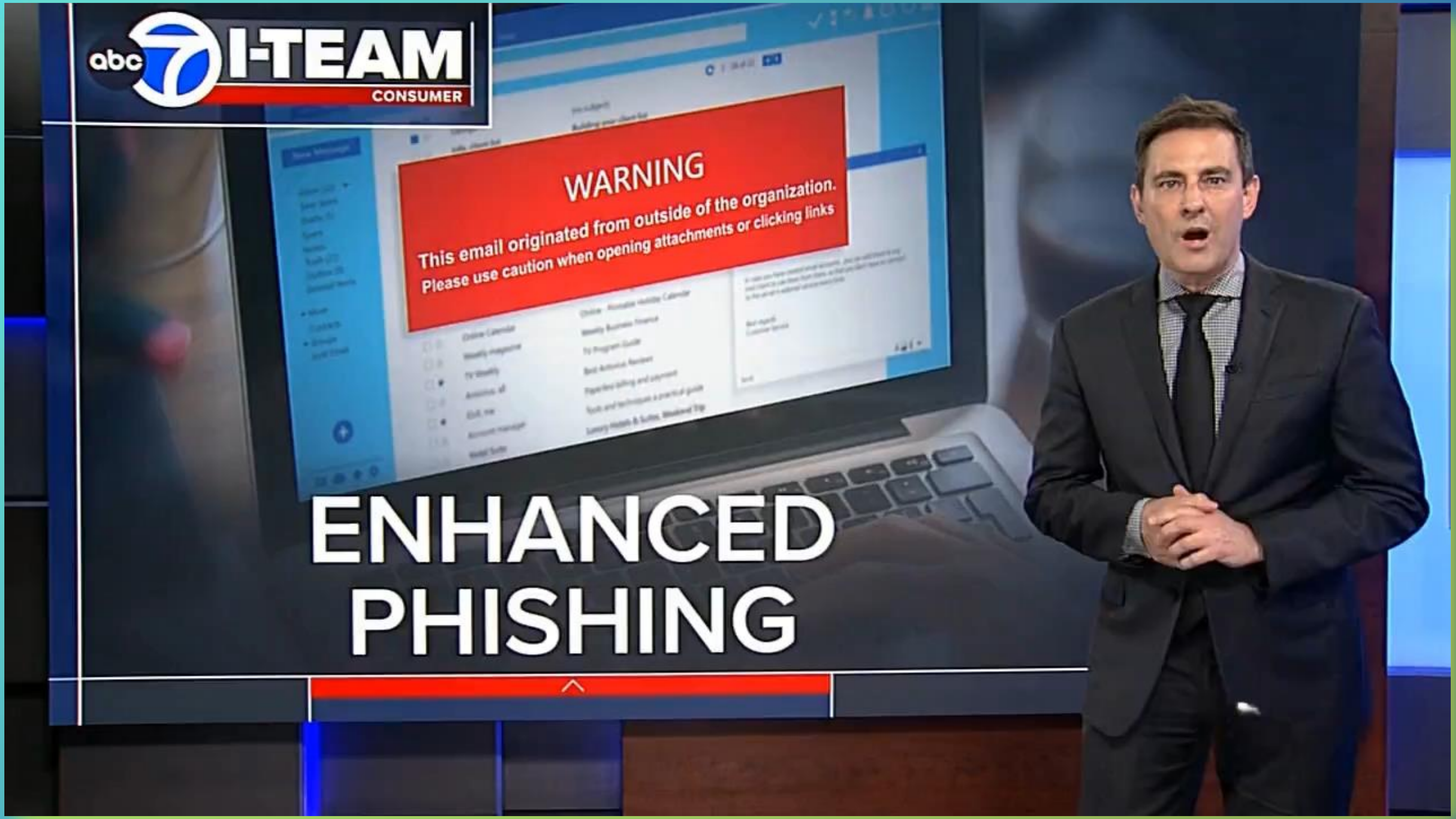


Nathan Poch
Masters in Cyber Security

Maurisa Dacosta
Masters in Computer Science

Rodrigo Arguello
Masters in Computer Science

Phishing in the Headlines: Why Smarter Detection Matters



Common Phishing Types

Phishing

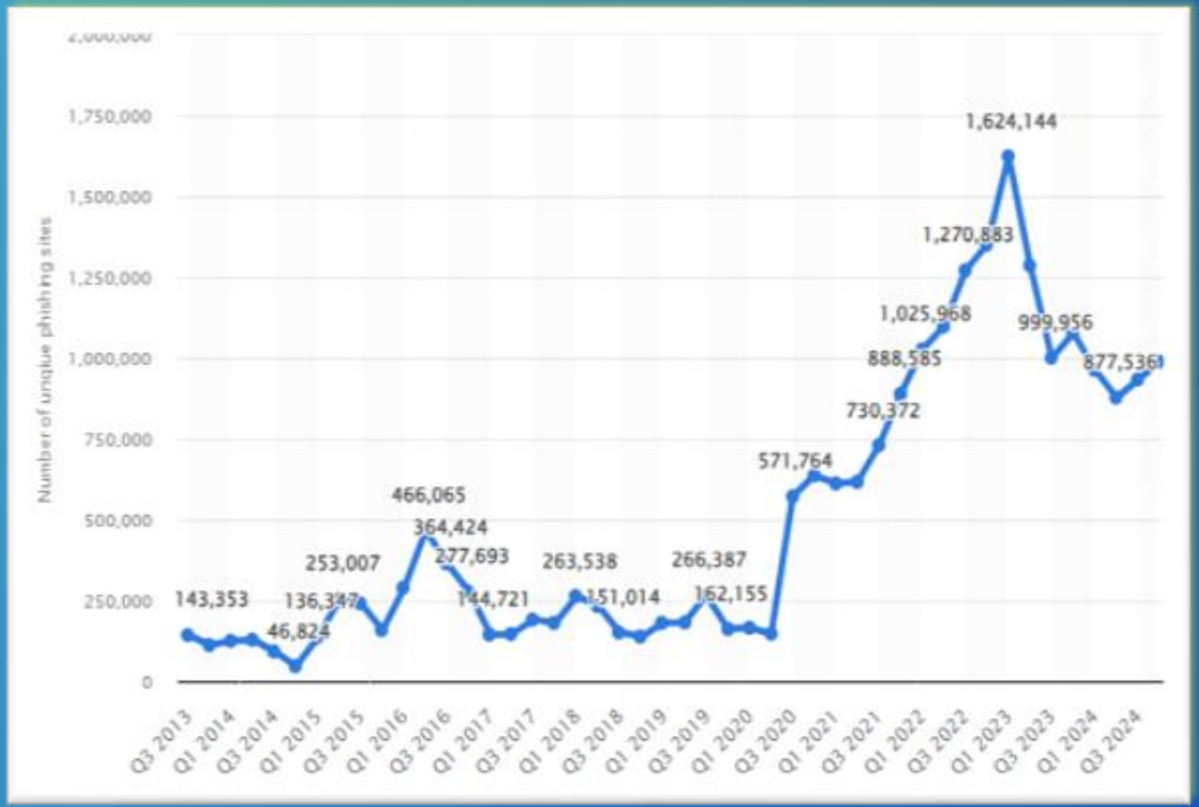
Uses **mass emails** to trick **individuals and groups** into revealing sensitive information.

Spear Phishing

Uses **personalized emails** to trick **individuals** into revealing information.

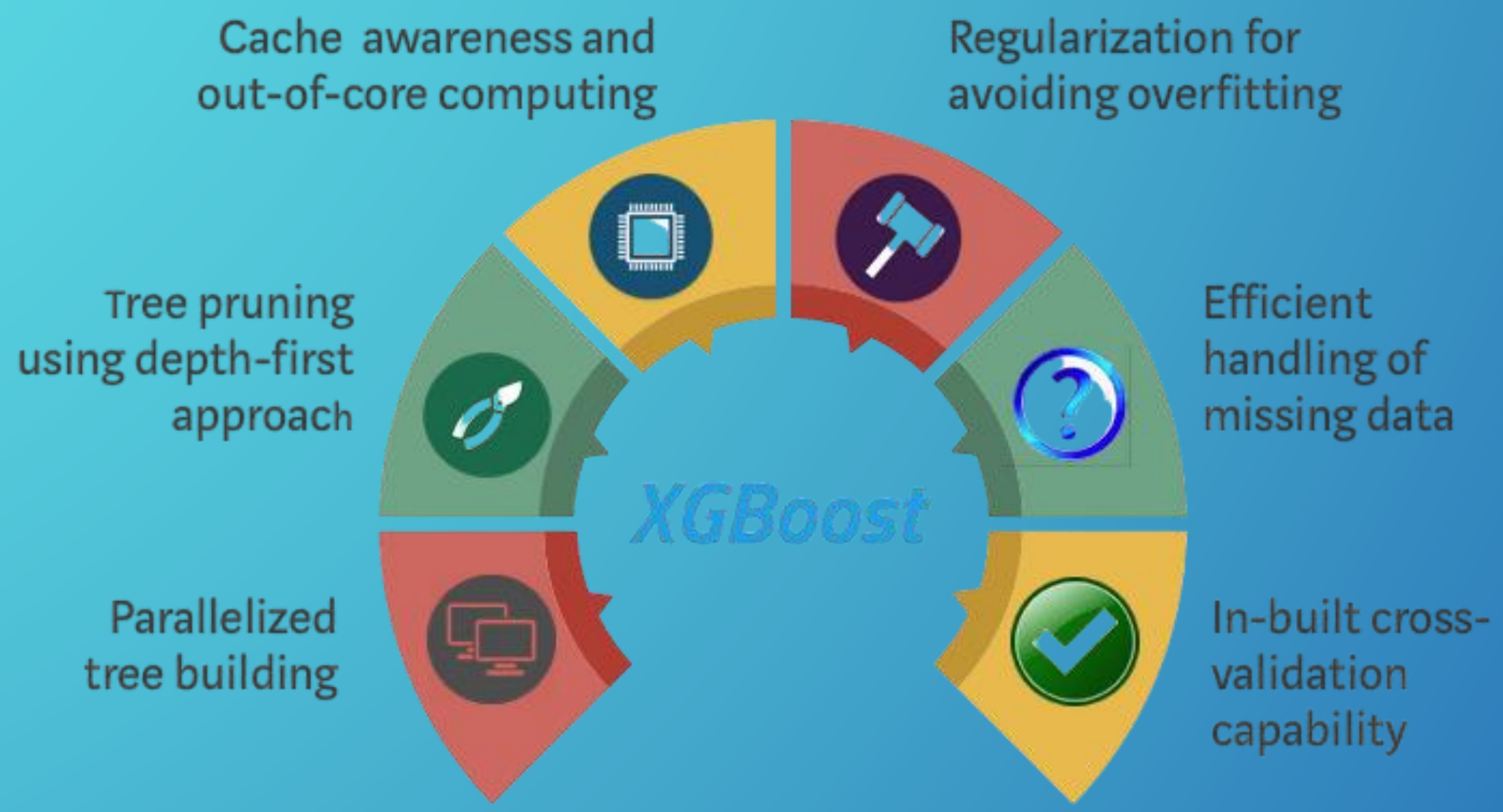
Whaling

Uses **personalized emails** to trick **high-value targets** into revealing information.



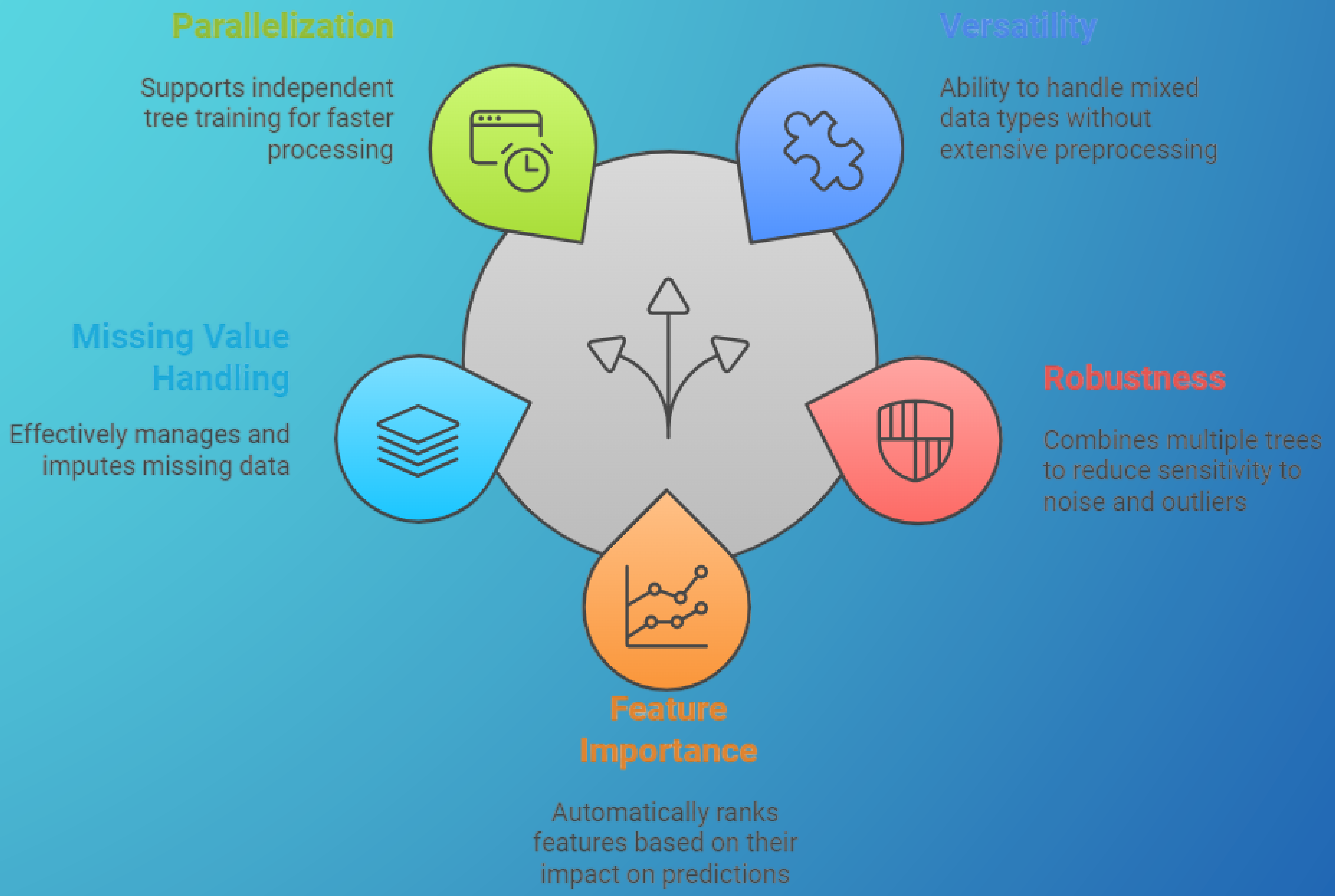
The number of detected phishing sites across the world in each quarter from the 3rd quarter of 2013 to the 4th quarter of 2024 (Statista, 2025).

Why XG Boost ?



Cache awareness and out-of-core computing
Regularization for avoiding overfitting
Tree pruning using depth-first approach
Efficient handling of missing data
XGBoost
Parallelized tree building
In-built cross-validation capability

Why Random Forest ?



Why MLP ?


Adapts to new threats
Learns non-linear patterns, detects anomalies
Supports real-time detection with low false positives
Provides quick predictions after training
Works well with large input data

Feature Extraction Using DistilBERT

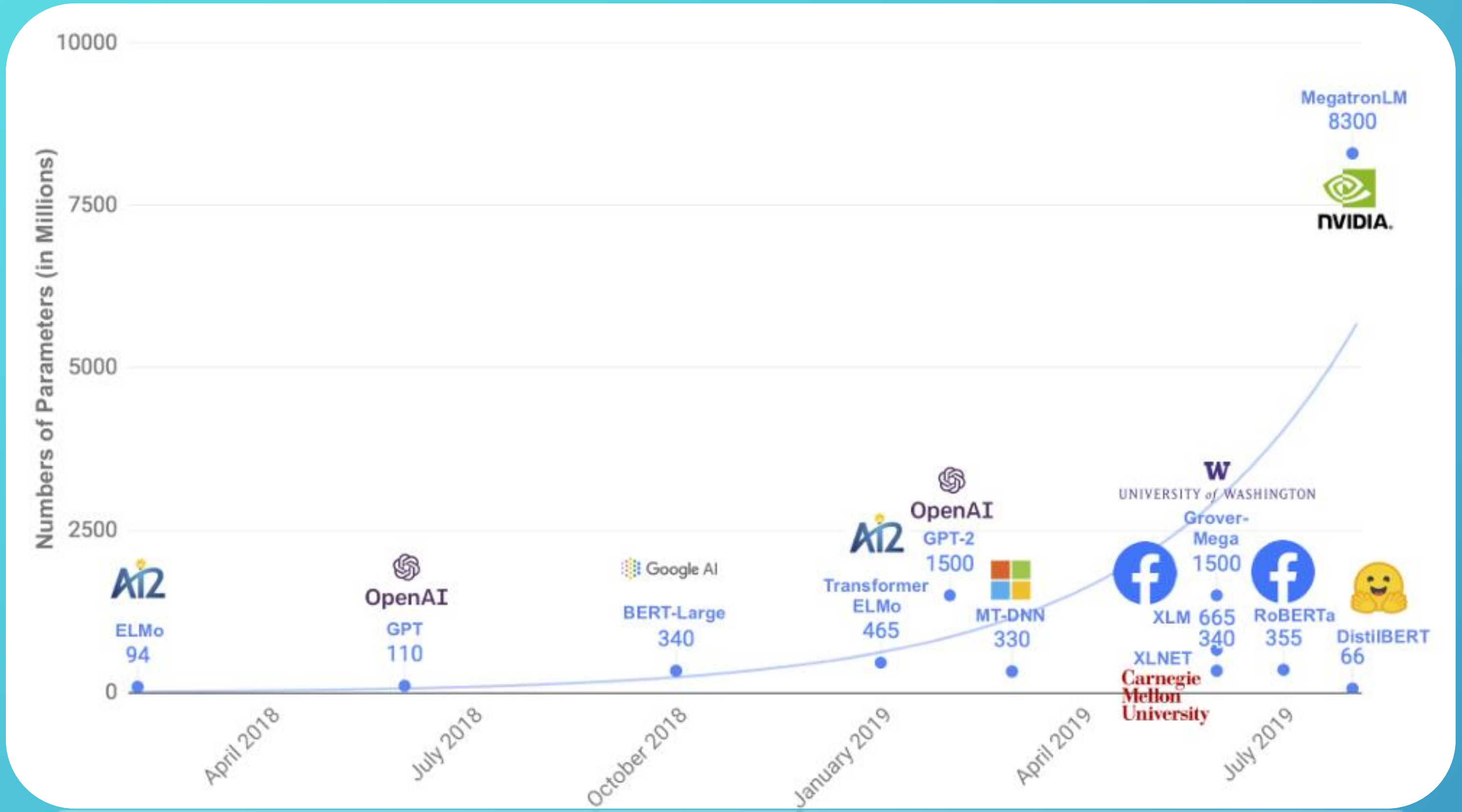


Table 1: **DistilBERT retains 97% of BERT performance.** Comparison on the dev sets of the GLUE benchmark. ELMo results as reported by the authors. BERT and DistilBERT results are the medians of 5 runs with different seeds.

Model	**Score**	CoLA	MNLI	MRPC	QNLI	QQP	RTE	SST-2	STS-B	WNLI
ELMo	68.7	44.1	68.6	76.6	71.1	86.2	53.4	91.5	70.4	56.3
BERT-base	79.5	56.3	86.7	88.6	91.8	89.6	69.3	92.7	89.0	53.5
DistilBERT	77.0	51.3	82.2	87.5	89.2	88.5	59.9	91.3	86.9	56.3

Table 2: **DistilBERT yields to comparable performance on downstream tasks.** Comparison on downstream tasks: IMDb (test accuracy) and SQuAD 1.1 (EM/F1 on dev set). D: with a second step of distillation during fine-tuning.

Model	IMDb (acc.)	SQuAD (EM/F1)
BERT-base	93.46	81.2/88.5
DistilBERT	92.82	77.7/85.8
DistilBERT (D)	-	79.1/86.9

Table 3: **DistilBERT is significantly smaller while being constantly faster.** Inference time of a full pass of GLUE task STS-B (sentiment analysis) on CPU with a batch size of 1.

Model	# param. (Millions)	Inf. time (seconds)
ELMo	180	895
BERT-base	110	668
DistilBERT	66	410

- ❖ 40% smaller, 60% faster than BERT
- ❖ 97% of BERT's performance
- ❖ Extracts semantic features from phishing emails/URLS

Evaluation Metrics

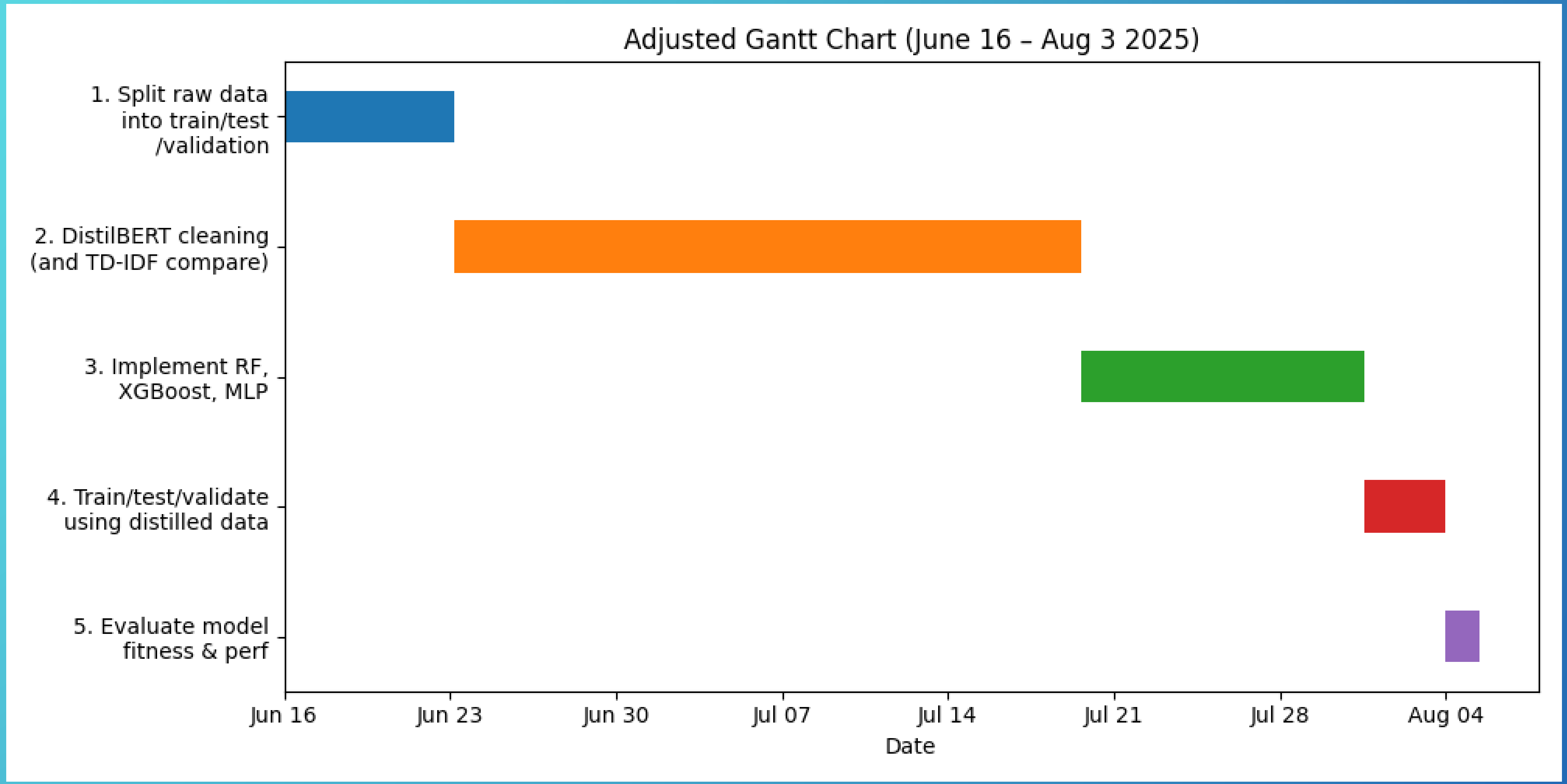
$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

$$\text{False Positive Rate (FPR)} = \frac{FP}{FP+TN}$$

$$\text{False Negative Rate (FNR)} = \frac{FN}{FN+TP} = 1 - Recall$$

Key: TP=True Positive, TN=True Negative, FP=False Positive, FN=False Negative

Timeline & Feasibility



References

Used in Presentation:

Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2019). DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. arXiv preprint arXiv:1910.01108.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019, June). Bert: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers) (pp. 4171-4186).

Owen, A., & White, E. (2024). Evaluating the Effectiveness of AI-Based Phishing Detection in IoT Communication Protocols.

Morde, V. (2019, June 23). *XGBoost algorithm: Long may she reign!* Medium. https://medium.com/data-science/https-medium-com-vishalmorde-xgboost-algorithm-long-she-may-rein-edd9f99be63d

SitePoint. (2025). Random Forest algorithm in machine learning. https://www.sitepoint.com/random-forest-algorithm-in-machine-learning/

Gu, J., & Xu, H. (2022, January). An ensemble method for phishing websites detection based on xgboost. In 2022 14th international conference on computer research and development (ICCRD) (pp. 214-219). IEEE.

REDDY, K. T. (2024) Enhancing Phishing Detection Through Multilayer Perceptron in Cybersecurity.

Shahrivari, V., Darabi, M. M., & Izadi, M. (2020). Phishing detection using machine learning techniques. arXiv preprint arXiv:2009.11116.

Potential Data Sets:

GitHub - rokibulroni/Phishing-Email-Dataset: A comprehensive dataset of phishing and legitimate emails curated for cybersecurity research and applications. This dataset is designed to help researchers, data scientists, and cybersecurity professionals develop, train, and evaluate models for phishing detection, email filtering, and threat analysis.

Tamal, M. (2023). *Phishing Detection Dataset* [Data set]. Mendeley Data. https://data.mendeley.com/datasets/6tm2d6sz7p/1