

UNIVERSITÀ DEGLI STUDI DI PADOVA



SCUOLA DI SCIENZE

CORSO DI LAUREA IN INFORMATICA

---

## Piano di lavoro

---

*Studente:*

Matteo Tiozzo - 2042882

*Azienda:*

Università di Padova, Dipartimento di  
Matematica



---

12 settembre 2024

## Contatti

**Studente:** Matteo Tiozzo, [matteo.tiozzo.1@studenti.unipd.it](mailto:matteo.tiozzo.1@studenti.unipd.it), + 39 3345263731

**Tutor aziendale:** Alessandro Galeazzi, [alessandro.galeazzi@unipd.it](mailto:alessandro.galeazzi@unipd.it),

**Azienda:** Università di Padova, Dipartimento di Matematica,  
Via Trieste, 63, 35131 Padova, <https://www.math.unipd.it/>

## Scopo dello stage

Lo scopo dello stage è acquisire competenze nell'applicazione di tecniche di Intelligenza Artificiale Interpretabile (Explainable AI) all'analisi e l'interpretazione delle caratteristiche delle Reti Generative Avversarie (GAN) nella generazione di malware. Lo stage si focalizzerà sull'esplorazione di metodi avanzati per migliorare la trasparenza e l'interpretabilità dei risultati prodotti dalle GAN, contribuendo allo sviluppo di soluzioni di cybersecurity interpretabili.

Lo studente dovrà acquisire competenze nell'utilizzo di strumenti di Explainable AI come Grad-CAM, Lime o tecniche di sensibilità all'occlusione, comprendere i meccanismi alla base delle GAN, e sviluppare la capacità di analizzare e comunicare i risultati attraverso report e visualizzazioni.

## Interazione tra studente e tutor aziendale

Gli incontri con il tutor aziendale Alessandro Galeazzi avverranno a cadenza settimanale. In tali occasioni lo studente discuterà dei progressi raggiunti, dei prossimi obiettivi, di eventuali aggiornamenti al piano di lavoro e di come migliorare la qualità del progetto.

## Prodotti attesi

Lo studente dovrà produrre una relazione scritta che illustri i seguenti punti:

- Revisione della letteratura esistente  
Effettuare una revisione della letteratura sulle tecniche che utilizzano le Reti Generative Avversarie (GAN) nel campo dell'analisi e rilevazione di malware.
- Raccolta del dataset di malware  
Raccogliere un dataset curato di eseguibili malware da fonti affidabili come Malwarebazaar, Virusshare, ecc., garantendo qualità e diversità dei dati. Utilizzare servizi come VirusTotal o AVClass2 per costruire un dataset di malware etichettato per tipologia.
- Conversione dei binari di malware  
Convertire i binari di malware in un formato adatto come input per le GAN
- Sviluppo del sistema di rilevazione malware  
Progettare e implementare un sistema di rilevazione del malware utilizzando algoritmi di deep learning come Convolutional Neural Networks (CNN), InceptionNet, XceptionNet e altri.

- **Addestramento dei modelli GAN**  
Sviluppare un'architettura GAN (es. DCGAN, WGAN) adatta alla generazione di malware. Monitorare metriche chiave come perdita, FID (Fréchet Inception Distance) e la qualità visiva dei campioni di malware generati.
- **Applicazione delle tecniche di Explainability**  
Analizzare le prestazioni tramite tecniche di Explainable AI come Grad-CAM e Lime.
- **Valutazione dell'interpretabilità (Analisi quantitativa) e Ablation analysis**  
Misurare la coerenza delle caratteristiche evidenziate su diversi campioni e tipologie di malware. Eseguire ablation analysis rimuovendo o modificando le caratteristiche chiave.

## Contenuti formativi previsti

Il progetto prevede che lo studente metta in pratica e approfondisca le sue conoscenze nell'ambito dell'Intelligenza Artificiale Interpretabile (Explainable AI) e delle Reti Generative Avversarie (GAN). Inizialmente, lo studente dovrà acquisire competenze nella comprensione e manipolazione di dataset di malware, con un focus particolare sull'analisi delle caratteristiche generate dalle GAN. Successivamente, è richiesto che familiarizzi con metodi per migliorare la trasparenza e l'interpretabilità dei modelli generativi, utilizzando tecniche come Grad-CAM e Lime. Durante l'attività di stage, lo studente potrà quindi approfondire le tecniche avanzate e gli strumenti utilizzati per rendere i modelli di intelligenza artificiale più trasparenti e interpretabili nel contesto della sicurezza informatica.

## Pianificazione del lavoro

### Pianificazione settimanale

- **Prima Settimana (40 ore)**
  - Revisione della letteratura e delle tecniche esistenti per le Reti Generative Avversarie
  - Identificazione e scaricamento di dataset di malware
  - Etichettatura dei dati per lo sviluppo del sistema di rilevamento
- **Seconda Settimana (40 ore)**
  - Pre-elaborazione e pulizia dei dati
  - Identificazione ed estrazione delle caratteristiche rilevanti
- **Terza Settimana (40 ore)**
  - Studio dei modelli di deep learning per identificazione di malware
  - Classificazione delle tipologie dei codici sorgenti
- **Quarta Settimana (40 ore)**
  - Sviluppo di esempi di reti avversarie utilizzando GAN
  - Creazione di un dataset per la valutazione del modello
- **Quinta Settimana (40 ore)**
  - Analisi delle caratteristiche del classificatore tramite Grad-CAM/Lime/Occlusion sensitivity
  - Analisi delle differenze tra malware originali e generati sinteticamente
- **Sesta Settimana (40 ore)**
  - Analisi della similarità delle caratteristiche dei malware nella stessa famiglia
  - Analisi della similarità delle caratteristiche dei malware tra famiglie diverse
- **Settima Settimana (40 ore)**
  - Descrizione e visualizzazione dei risultati
  - Confronto con i ricercatori coinvolti per discutere i risultati ottenuti
- **Ottava Settimana (20 ore)**
  - Applicazione dei feedback ricevuti
  - Redazione documentazione e relazione finale;



## Ripartizione ore

La pianificazione, in termini di quantità di ore di lavoro, sarà così distribuita:

Durata in ore	Descrizione dell'attività
<b>80</b>	<b>Background</b>
20	<i>Revisione delle tecniche e della letteratura</i>
60	<i>Creazione del dataset</i>
<b>140</b>	<b>Progettazione e Creazione dei modelli</b>
50	<i>Selezione del modello di malware detection</i>
50	<i>Selezione del modello di malware detection</i>
40	<i>Addestramento dei modelli</i>
<b>80</b>	<b>Valutazione dei Modelli</b>
50	<i>Applicazione delle tecniche di Explenability</i>
20	<i>Analisi dei risultati</i>
10	<i>Stesura documentazione finale</i>
<b>Totale ore</b>	<b>300</b>

## Obiettivi

### Notazione

Si farà riferimento ai requisiti secondo le seguenti notazioni:

- $O$  per i requisiti obbligatori, vincolanti in quanto obiettivo primario richiesto dal committente;
- $D$  per i requisiti desiderabili, non vincolanti o strettamente necessari, ma dal riconoscibile valore aggiunto;
- $F$  per i requisiti facoltativi, rappresentanti valore aggiunto non strettamente competitivo.

Le sigle precedentemente indicate saranno seguite da una coppia sequenziale di numeri, identificativo del requisito.

### Obiettivi fissati

Si prevede lo svolgimento dei seguenti obiettivi:

- Obbligatori
  - O01: Creazione di un dataset di malware classificato per categoria
  - O02: Implementazione di un sistema per identificare i malware basato su deep learning
  - O03: Creazione di un modello avversario (GAN) per generare malware
  - O04: Valutazione delle performance del modello secondo tecniche di Explainable AI
  - O05: Valutazione delle performance del modello tramite ablation analysis
- Desiderabili
  - D01: Analisi esaustiva dei modelli tramite Explainable AI
  - D02: Analisi approfondita dei risultati degli esperimenti;
- Facoltativi
  - F01: Implementazione di diversi modelli per identificazione di malware.
  - F02: Implementazione di diverse architetture per la generazione di malware.



---

## Approvazione

Il presente piano di lavoro è stato approvato dai seguenti

Alessandro Galeazzi

Tutor aziendale

Matteo Tiozzo

Stagista

Prof. Alessandro Brighente

Tutor interno

Data