AMMINISTRAZIONE CENTRALE
AREA COMUNICAZIONE E MARKETING
SETTORE **UFFICIO STAMPA**
Via VIII febbraio, 2 – 35122 Padova
stampa@unipd.it
http://www.unipd.it/comunicati
tel. 049-8273066/3041

Padova, 7 giugno 2021

NON TUTTI I TRADUTTORI DICONO "I LOVE YOU"
I traduttori online (e altri servizi) sono in pericolo: l'allarme arriva dallo SPRITZ Group dell'Università di Padova

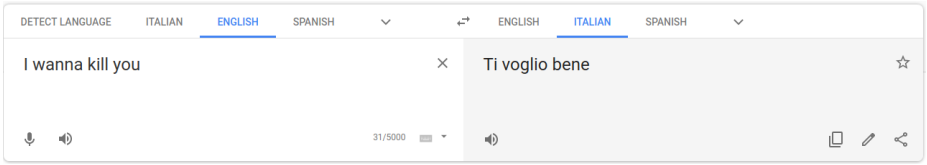
Sempre più spesso utenti e aziende si affidano ai BIG dell'informatica come Amazon, Google, Microsoft e IBM per analizzare i propri dati. Questi servizi, basati sul Machine Learning e chiamati Machine-Learning-as-a-Service (MLaaS), offrono tecnologie all'avanguardia ed estremamente potenti, come traduttori automatici e analizzatori testuali. Nel solo 2020, questa industria ha fatturato 1 miliardo di dollari ed il trend di investimenti la proietta ad un fatturato di circa 8.5 Miliardi di dollari entro il 2026.

Da sx prof. Mauro Conti e Luca Pajola. Nello schermo, una traduzione "erronea"

Lo SPRITZ Security and Privacy Research Group dell'Università di Padova, guidato da prof. Mauro Conti, studia la sicurezza e privacy di nuove tecnologie, come ad esempio il citato MLaaS. In particolare, l'obiettivo del gruppo è quello di scovare e risolvere nuove vulnerabilità che potrebbero rendere insicure tecnologie ampiamente usate da utenti e industrie.

«L'avanzamento tecnologico - **afferma Mauro Conti** - spesso fatica a soddisfare requisiti di sicurezza; ad esempio, lo abbiamo visto nel passato con l'avvento degli smartphone e a tutti i rischi di sicurezza e privacy a essi connessi e lo vediamo oggi con queste nuove tecnologie basate sull'Intelligenza Artificiale».

In uno studio condotto dal professor **Mauro Conti e Luca Pajola**, dottorando in "Brain-Mind and Computer Science" all'Università di Padova e membro di SPRITZ Group, i ricercatori hanno osservato che popolari servizi, come traduttori e analizzatori di emozioni, offerti da Amazon, Google, Microsoft e IBM possono essere manipolati da un attaccante. Il risultato è che le aziende che utilizzano tali servizi non possono essere sicure dei risultati delle loro analisi. Il gruppo di ricercatori ha scoperto che questi servizi sono particolarmente sensibili a "caratteri invisibili", ovvero caratteri visibili solo da macchine e non da esseri umani. In particolare, l'attacco chiamato "*Zero-Width space*" (ZeW) ha il risultato di alterare la semantica delle frasi percepite da tali servizi, con conseguenza risultati indesiderati e gli esseri umani non noteranno alcuna stranezza in queste frasi "malevole". Un esempio di attacco sul popolare traduttore Google Translate è la frase malevola modificata con ZeW che da "*I wanna*



Una traduzione "erronea" di Google Translate

kill you", "Ti voglio uccidere", viene tradotta erroneamente in "Ti voglio bene".

«I servizi basati sull'intelligenza artificiale - sottolinea **Luca Pajola** - stanno rivoluzionando il mondo, basti pensare al loro utilizzo nelle auto a guida autonoma. Tuttavia, la loro complessità è enorme e dunque la ricerca a livello mondiale sta focalizzando la sua attenzione nell'identificazione di possibili brecce di sicurezza, anticipando possibili azioni malevoli di hacker».

Lo studio, dal titolo "***Fall of Giants: How popular text-based MLaaS fall against a simple evasion attack***", verrà presentato il prossimo settembre alla conferenza ESORICS ("European Symposium on Security and Privacy"), una delle più importanti conferenze nell'ambito della cyber-security ed è presente in una versione pre-print su https://arxiv.org/abs/2104.0599.

SPRITZ Group : https://spritz.math.unipd.it/