CENTER FOR INEQUALITY AND OPEN SOCIETY

Platební podvody v éře generativní a agentní AI: jaké výzvy čekají rozhodovací praxi?

Případová studie České republiky a Německa

Mgr. Bc. Anežka Karpjáková
Ústav práva a technologií, PrF MU

Spolufinancováno Evropskou unií

MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

UNIVERZITA KARLOVA

MUNI PRÁVNICKÁ FAKULTA

VŠE VYSOKÁ ŠKOLA EKONOMICKÁ V PRAZE

Téma příspěvku

- **Jak mění generativní a agentní AI platební podvody?**
- **Právní výzva: Jak budou v případech sofictikovaných podvodů využívajících generativní a agentní AI soudy pravděpodobně hodnotit hrubou nedbalost klienta? Případová studie české a německé rozhodovací praxe**

Současná podoba platebních podvodů

Sociální inženýrství = nejrozšířenější technika útoku na bankovní klienty (zejména phishing, vishing, smishing)

Cíl útočníka: manipulací přimět klienta k jednání, které umožní platbu na účet útočníka, např.:

- sdělení přihlašovacích údajů nebo verifikačního kódu
- aktivace nového mobilního zařízení, digitální karty nebo digitální peněženky (např. Apple Pay)
- vzdálený přístup k zařízení klienta
- provedení transakce samotným klientem

Různorodé scénáře: falešní bankéři, falešné autority a organizace, blízké osoby…

Zvyšuje se nejen počet útoků, ale i kvalita jejich provedení – značný vliv **generativní a agentní AI**.

4,2 mld. EUR

podvodné platební transakce v EHP za rok 2024, meziročně +17 %

EBA / ECB, 2025 Report on Payment Fraud

+37 % meziroční nárůst zasažených bankovních klientů v ČR

– téměř 50 000 za 1. pololetí 2024

(Česká bankovní asociace)

Generativní a agentní AI: co mění v podvodech

Generativní AI

Počátkem roku 2025 využívalo gen AI přes 80 % pozorovaných phishingových kampaní (ENISA).

- **Generování přesvědčivého obsahu** – zejména kvalitní e-maily, zprávy (odstranění gramatických chyb, nepřirozených formulací, personalizace), klonování hlasu a vedení konverzaci v reálném čase, generování videí a vedení videohovorů
- **Největší hrozba = deepfake (Europol):**
  - syntetické klonovaní hlasu, fotografie nebo video skutečné osoby (klon hlasu z několika sekund, video z jedné fotografie),
  - vydávání se za blízké osoby
  - 35–50 % lidí nerozezná deepfake video

Agentní AI

Na základě stanoveného cíle funguje autonomně – plánuje kroky, používá nástroje, prohlíží internet a v reálném čase reaguje na oběť. Slouží k různým účelům:

- **Personalizace** – AI zanalyzuje digitální stopu oběti a vytváří útok na míru (spear phishing)
- **Řízení kampaně** – automatizovaně vyhledá oběť, shromáždí o ní informace z veřejných zdrojů, osloví ji a vyhodnotí reakci, schopnost improvizovat.
- **Škálování** – vytváří velké množství variant zpráv, které lze rozesílat pomocí botnetů, a současně vede více konverzací

Co obě mění pro oběť: mizí varovné signály (gramatické chyby, pevný scénář, neobratná odpověď); útok je personalizovaný; hlas i tvář blízké osoby; tisíce obětí bez call centra; nástroje dostupné a levné → **těžší rozpoznatelnost pro potenciální oběť**

Klasické podvody vs. podvod s deepfake

Klasický phishing a vishing

PHISHINGOVÝ E-MAIL

MB **Moje Banka | Klientské centrum**
klientske.centrum@mojebanka-online.cz
dnes 09:14

Aktualizace bezpečnostních údajů k Vašemu účtu

Vážený kliente, v rámci aktualizace bezpečnostních prvků Vás žádáme o ověření přihlašovacích údajů. Provedte je prosím do 3 dnů, jinak může být přístup omezen.

Ověřit údaje https://mojebanka-online.cz/overeni

VISHINGOVÝ HOVOR

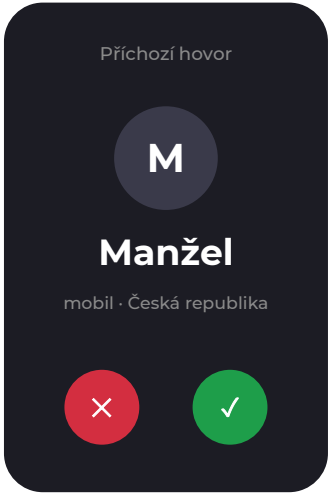
B **Moje Banka – bezpečnostní oddělení**
800 123 456 · zobrazené číslo banky

„Dobrý den, zaznamenali jsme podezřelou platbu z Vašeho účtu. Abychom ji zablokovali, nadiktujte mi prosím kód z SMS, kterou Vám právě posíláme.“

SMS · Moje Banka
Kód 483 921 slouží k aktivaci nového zařízení. Banka ho nikdy nevyžaduje – nikomu ho nesdělujte.

- e-mail: obecné oslovení, jazyková chyba, doména a odkaz mimo banku
- hovor: neznámý hlas žádá kód – SMS říká, k čemu slouží, a varuje
- Generické informace
- Mediálně známé průběhy podvodů

Podvod s využitím deepfake



„Ahoj, jsem u účetní a potřebuju se z mobilu dostat do našeho účtu, ale telefon mi hlásí, že si musím znovu aktivovat aplikaci. Přijde ti teď do tvojí aplikace žádost o potvrzení nového zařízení – jen to prosím odklepni, potřebuju to hned…“

- telefonní číslo odpovídá dané osobě (spoofing)
- klon hlasu dané osoby a znalost detailů
- kód z SMS aktivuje cizí zařízení – platbu pak zadá útočník sám

CENTER FOR INEQUALITY AND OPEN SOCIETY

Výzva pro soudy: Kdo nese ztrátu z podvodné platby?

Nejčastější otázka, kterou soudy a mimosoudní orgány u těchto podvodů řeší.

Pachatele se v drtivé většině případů nepodaří zjistit → **uplatní se odpovědnostní rámec pro neautorizované transakce: ztrátu nese buď poskytovatel platebních služeb (banka), nebo jeho klient.**

Právní úprava: Odpovědnostní rámec pro neautorizované platební transakce (rozdělení odpovědnosti mezi PSP a klienta)

Neautorizované transakce = transakce, ke kterým klient nedal souhlas (čl. 64 odst. 1 PSD2).

Souhlasem se rozumí samotné technické provedení: kdo transakci v internetovém bankovnictví skutečně zadal.

Neautorizovaná transakce

zadal ji útočník

Např. klient omylem povolí nové zařízení s mobilní aplikací banky – a útočník z něj transakci provede sám.

Autorizovaná transakce

zadal ji sám klient

Např. investiční podvod – útočník přesvědčí klienta, aby mu sám poslal peníze v domnění, že je investuje.

Odpovědnostní rámec pro neautorizované transakce

Rámec vyplývá z unijní směrnice PSD2 a stanoví, kdo ztrátu z neautorizované transakce ponese.

PRAVIDLO

Odpovědnost nese primárně banka

Poskytovatel platebních služeb klientovi neprodleně vrátí částku neautorizované transakce.

čl. 73 PSD2

NEJDŮLEŽITĚJŠÍ VÝJIMKA

Odpovědnost nese v plném rozsahu klient

pokud z **HRUBÉ NEDBALOSTI** porušil některou ze svých smluvních povinností – například:

- sdělil citlivé údaje (přihlašovací údaje, kódy) cizí osobě
- umožnil cizí osobě přístup ke svému bankovnímu účtu

čl. 74 odst. 1 ve spojení s čl. 69 PSD2

- **Hrubou nedbalost PSD2 nedefinuje** – recitál 72: „více než pouhá nedbalost, jednání vykazující značný stupeň neopatrnosti“
- Důkazy, kritéria pro určení stupně nedbalosti a hranice mezi prostou a hrubou nedbalostí se posuzují **podle vnitrostátního práva členských států.**
- Dle EBA: **rozložení ztrát se mezi členskými státy výrazně liší** – i kvůli odlišným výkladům hrubé nedbalosti.

Případová studie: ČR a Německo

Doposud neexistují pravomocná rozhodnutí, která by řešila odpovědnost za neautorizované platební transakce v případech, kdy byl k provedení podvodu využit deepfake nebo jiný nástroj generativní AI.

Výzkumné otázky

1 Jaká kritéria pro určení hrubé nedbalosti se v rozhodovací praxi v Česku a Německu v současné době aplikují a jak by obstála u podvodů s využitím deepfake a dalších nástrojů generativní či agentní AI?

2 Jak hodnotící rámec zpřesnit (de lege ferenda)?

Vzorek

- **Německo: 49 soudních rozhodnutí** – zemské, vrchní zemské a obvodní soudy, Spolkový soudní dvůr
- **Česko: 14 nálezů finančního arbitra** – české soudy zatím žádné zveřejněné rozhodnutí

Shodná kritéria výběru: 1) případ řeší hrubou nedbalost v rámci neautorizované transakce způsobené phishingem, vishingem či obdobným podvodem, 2) klient byl spotřebitel, 3) rozhodnutí bylo vydáno v období 11/2009–02/2026

Německo: sofistikovanost útoku

Celkem hrubá nedbalost ve 33 ze 49 rozhodnutí

Málo sofistikované – **zjevné anomálie, nesrovnalosti rozpoznatelné bez odborných znalostí** (gramatické chyby, absurdní obsah, desítky kódů)

Středně sofistikované – **věrohodné a přesvědčivé, avšak obsahující rozpoznatelné varovné signály** (převážně případy falešných bankéřů: vishing se spoofingem tel. čísla banky, znalost jména osobní bankéřky); rozhodovalo ignorování varování v aplikaci či SMS a nevyužitý čas ověřit si situaci

Vysoce sofistikované – **průměrný uživatel měl omezenou možnost podvod odhalit a novost podvodného scénáře/absence varování ze strany banky.** Např. bankovní trojan (SpyEye, tzv. man-in-the-browser), aktivace Apple Pay nebo nového zařízení s neurčitou notifikací v autorizační aplikaci, SIM swap.

Čím hůře se dal podvod odhalit, tím je menší pravděpodobnost hrubé nedbalosti.

Německo: identifikovaná kritéria

Subjektivní přístup pro posuzování hrubé nedbalosti = soud zkoumá i individuální vlastnosti klienta (např. věk, zkušenosti, digitální gramotnost, profese). Pro určení hrubé nedbalosti tak musí být jednání **objektivně závažné** a **subjektivně neomluvitelné** (BGH, XI ZR 107/24).

Objektivní faktory	n
PROTI KLIENTOVI	
Vyzrazení kódu či aktivačního odkazu třetí osobě	28
Obecné povědomí veřejnosti o phishingu	27
Zjevná nelogičnost požadavku	18
Opakované vyzrazení nebo potvrzení bezpečnostních prvků	15
Ignorování údajů v autorizační aplikaci	13
Kliknutí na podvodný odkaz v e-mailu	11
Neověření pravosti komunikace u banky	10
VE PROSPĚCH KLIENTA	
Absence upozornění banky na konkrétní podvodné scénáře	10
Spoluzavinění banky (neurčitý text SMS / notifikace)	5
Novost scénáře podvodu / nízké veřejné povědomí o scénáři	4

Subjektivní faktory	n
PROTI KLIENTOVI	
Délka a zkušenost s internetovým bankovnictvím	32
Vědomé přehlížení rizika	16
Profesní zázemí (IT, právo, bankovnictví)	10
Dostatek času na rozmyšlení	9
Inteligence	3
VE PROSPĚCH KLIENTA	
Stres a psychický tlak	12
Zranitelnost / zvláštní osobní okolnosti (např. jazyková bariéra)	8
Augenblicksversagen – okamžité selhání (přijato 3×)	5

Faktory, které soud nezohlednil: míru personalizace útoku; osobu, za kterou se útočník vydává.

ČR: analýza nálezů FA

- **14 nálezů: ve všech hrubá nedbalost hledána**
- **Průběh podvodů:** ve většině případů se jednalo o méně sofistikovaný průběh podvodu (zejména phishingové e-maily s falešným odkazem na internetové bankovnictví)
- **Přístup arbitra:** téměř výlučně **objektivní a kumulativní**
- **Argumentace: porušení smluvních podmínek a kumulace pochybení (ve všech 14 případech):** kliknutí na odkaz v phishingovém e-mailu (8), zadání přihlašovacích nebo bankovních údajů na podvodném webu (12), vyzrazení verifikačního kódu třetí osobě (11), neověření pravosti webu (9), ignorování varovných notifikací banky (9), pozdní nahlášení podvodné transakce bance (4).

14 / 14

hrubá nedbalost klienta shledána ve všech nálezech

Jak by dle stávajících kritérií byly deepfake podvody posouzeny?

Německo:

- určité případy s využitím věrohodného deepfaku by pravděpodobně spadaly do kategorie vysoce sofistikovaných scénářů (novost podvodů, omezená schopnost průměrné osoby deepfake rozeznat) → ve prospěch klienta.
- Vždy však bude záležet na objektivních a subjektivních faktorech konkrétního případu.

ČR: praxe arbitra bez výkladových vodítek → u technologicky pokročilých podvodů **obtížně předvídatelná**

Návrh de lege ferenda

Rámec nenahrazovat, ale zpřesnit o faktory, které praxe dosud opomíjela.

1. **Míra personalizace útoku**
2. **Kvalita obsahu vytvořeného AI**
3. **Vydávání se za osobu, kterou klient zná**
4. **Konkrétní varování banky před daným způsobem útoku**

Příklad ze slidu č. 5 (telefonát od manžela)

Personalizace – vysoká, zná události minulého týdne
Kvalita obsahu – klon hlasu nerozeznatelný
Známá osoba – vlastní manžel
Varování banky – před klonovaným hlasem blízké osoby v Německu časté, v ČR spíše výjimečné.

Děkuji za pozornost

Tato prezentace byla podpořena Evropským fondem pro regionální rozvoj v rámci projektu „Centrum pro nerovnost a otevřenou společnost" (č.: CZ.02.01.01/00/23_025/0008690).