

6

Máte dvacet vteřin na splnění podmínek

Neschopnost technologického průmyslu řešit škodlivé používání produktů se stala celosvětovou ostudou. Sociální sítě zaneřádili trollové a tyrani; do našich soukromých digitálních životů nevítaně pronikly firmy a vlády. Rozvíjející se technologie bohužel nabízejí nové možnosti útoků. V osobních vztazích lidí s nerovnou technologickou gramotností poskytuje ovládání domácí sítě jednomu partnerovi tichou moc: zhrzení milenci se mohou zlomyslně mstít prostřednictvím inteligentní domácnosti a děsit bývalé partnerky neočekávaným rozsvěcením světel a zvoněním zvonku nebo je špehovat domácími kamerami.¹⁴⁸ Mezitím naše představy o osobním prostoru zkresluje virtuální realita. Virtuální osahávání se vzhledem k působivému pocitu fyzické přítomnosti a nemožnosti vynutit si osobní hranice může zdát až příliš reálné.

Hackeri a podfukáři budou mít velký zájem na tom, aby o nic nepřišli. Připojené automobily se stanou lákavými cíli pro požadování výkupného – autorizujte tuto šifrovanou transakci, nebo vám zablokují brzdy – a digitální

¹⁴⁸ Nellie Bowles, „Thermostats, Locks and Lights: Digital Tools of Domestic Abuse“, *The New York Times*, 23. června 2018, nytimes.com.

asistenti by se kvůli (skutečné i metaforické) blízkosti uživatelům mohli nevědomky stát spiklenci sociálního inženýrství.

Představte si malwarovou aplikaci, která zachytí vaše požadavky virtuální asistentce Cortaně a jejím hlasem požádá uživatele o nějakou informaci. [...] Tato imitace rázem získá přístup k osobním údajům, které může využít v dalších fázích krádeže identity. [...] Budeme potřebovat ověřitelný způsob pro autentizaci agentů. – Chris Noessel¹⁴⁹

Nad průmyslem zabezpečení jako Damoklův meč visí kvantové počítače. Kvantové procesory vyřeší matematické problémy, které dnes považujeme za neřešitelné, a mohou případně rozbít běžnou ochranu online plateb, blockchainů a peněženek s kryptoměnou. Vytvoříte-li životaschopný kvantový procesor a budete-li dost chytří na to, abyste ho utajili, můžete ukrást celý svět. Podle většiny odhadů tato kvantová nadvláda nastane až za několik let: výroba kvantových čipů je d'ábelsky složitá a nákladná – máme tedy čas na to, abychom svá šifrovací opatření modernizovali. Nějaký tajemný syndikát už nicméně někde na kvantový poklad číhá.

Pro snižování rizik zneužití produktu existují zavedené postupy – funkce blokování/ztišení, zákazy, vymazání, mechanismy pro pozitivní/negativní hodnocení a omezení opětovného vstupu („Váš přístup je pozastaven, dokud neodstraníte toto“) –, po algoritmické páce se však podle očekávání natahují i technologické firmy. Je jistě lepší mít algoritmy, které lze škálovat, než muset zaměstnávat tisíce moderátorů obsahu, ovšem samotné technologie problém vyřešit nemohou. Jigsaw's Perspective, dceřiná společnost projektu Alphabet, prý rozpozná 92 % toxických komentářů, přičemž počet komentářů falešně pozitivních dosahuje překvapivě nízkého počtu 10 %, algoritmický zásah však bude vždy náchylný ke známé předpojatosti. Blokováním negativních postojů se omezí také smysluplný obsah, zatímco precizně koncipované projevy nenávisti sítí proklouznou. Jednotlivých „14 slov“ nechvalně známého sloganu o nadřazenosti bílé rasy je neškodných: skutečnou hrozbou je jejich spojení a použití v některých kontextech.¹⁵⁰

¹⁴⁹ Chris Noessel, *Designing Agentive Technology* (Rosenfeld Media, 2017).

¹⁵⁰ Caroline Sinders, „Silicon Valley's harassment problem“, *VICE News*, 28. května 2017, news.vice.com. [Tzv. „14 slov“ – „We must secure the existence of our people and

V ideálním případě je lepší zneužívání bránit než je řešit; újmě je lepší předcházet než se s ní vypořádávat až následně. Ve skutečnosti však problém potřebuje prevenci i nápravu. Hovořili jsme již o taktice při procesu návrhu, například o určených oponentech a personách non grata; měli bychom však vycházet i z hlubokého porozumění souvislostem a důsledkům zneužití. Výzkumné týmy by měly na svých platformách hovořit s (minulými, současnými i potenciálními) oběťmi zneužívání; podobně jako ve skutečnosti totiž toto zneužívání pravděpodobně neúměrně postihuje ženy a osoby jiné barvy pleti než bílé.

Lidé, které lze viditelně identifikovat jako osoby marginalizovaného rasového původu či genderu, mají s internetem zcela jiné zkušenosti. Zdá se jim nevrživější a více je vyčerpává, a stává se ještě nevrživějším a vyčerpává je ještě víc, pokud tito lidé patří do vícero takových skupin. – Sarah Jeongová¹⁵¹

Výzkum bude nepochybně účinnější, pokud týmy učiní (bohužel příliš vzácný) krok směrem k veřejnému sdílení vlastních poznatků; třebaže se zneužívání může na každé platformě projevat jinak, základní vzorce útoků a jejich dopady na uživatele se jistě překrývají. Silicon Valley toto obtěžování nejlépe vyřeší prostřednictvím spolupráce.

Pokusy řešit zneužívání, jež nemají oporu v řádném výzkumu, mohou dopadnout velmi špatně. Sdílením poznatků shromážděných od cílových skupin, či dokonce přizváním těchto osob do designového studia se riziko zhoršení situace sníží. Různorodý interní tým bude také se záplem hledat potenciální chyby v zásadách boje proti zneužívání a pravděpodobně porozumí dopadům na širší skupinu uživatelů: bude to taková etická infrastruktura v praxi. Vědci však musí být ve střehu. Jakmile se totiž pachatelé obtěžování dozvědí, že firma naslouchá obětem, často tvrdí, že skutečně poškozenou stranou jsou oni. Dynamika zneužívání může být složitá. Některé projevy obtěžování vyvolávají prudké protesty a eskalují; někdy se

a future for white children" (Našim lidem musíme zajistit existenci a bílým dětem budoucnost) – formuloval zakládající člen teroristické organizace The Order (Řád) a odsouzený zločinec David Lane (1938–2007). Pozn. překl.]

¹⁵¹ Sarah Jeong, *The Internet of Garbage* (Forbes Media, 2015).

může obtěžování maskovat a zvenčí působit jen jako další online šarvátka. Užitečnými vodítky mohou být základní hodnoty a osobní ctnosti: při analýze hádky věnujte pozornost nejen tomu, jak začala, ale i tomu, kdo ztělesňuje hodnoty, na nichž záleží; v případě pochybností upřednostněte právě jeho potřeby.

Moderování a svoboda projevu

Výzkum by měl poté vést k rázné politické reakci. Společnosti provozující sociální sítě – zejména Twitter a Reddit – čekaly na vytvoření a prosazení řádných pravidel chování příliš dlouho a začaly jednat, až když se zneužívání vymklo kontrole. Dnes zjišťují, že zpříšňování laxní politiky po příliš mnoha letech dá hrozně moc práce. Promyšlená, včas zavedená pravidla jsou podpůrnou konstrukcí vznikající kultury uživatelů. Žádná velká komunita se neohlídá sama, ovšem pokud lidé přijmou rozumné zásady, jež budou předpokladem členství v prosperující komunitě, často je budou prosazovat za vás.

Zásady potřebují také boti. Zvědavci i nadřzenci moc rádi testují hranice digitálních agentů; dnes jsou běžně sexuálně obtěžováni i konverzační boti. Vzhledem k tomu, že tři ze čtyř hlavních hlasových asistentů – Alexa, Cortana a Siri – jsou ženského rodu, nikoho to zřejmě nepřekvapí; do robotů si lidé podle všeho bez váhání projektují rasové a sexuální stereotypy.¹⁵² Dnes se boti po sexuálních poznámkách vzmožou jen na mírný odpor, pravděpodobně proto, aby nepůsobili konfrontačně či prudérně.¹⁵³ Ochrana botů před zneužíváním je však morální čin, který znamená přínos i pro širší společnost. Brzy se boti stanou agenty augmentace a v oblasti služeb budou působit společně s lidským personálem; pokud tedy chrlíte nadávky na digitálního asistenta, brzy začnete zneužívat i člověka. Rozhodné a spravedlivé odmítnutí či odklonění přesvědčí uživatele, že legrace, o kterou mu jde, se konat nebude, stanoví hranice toho, jak by se lidé na vaší platformě

¹⁵² Megan Strait et al, „Robots Racialized in the Likeness of Marginalized Social Identities are Subject to Greater Dehumanization than those racialized as White“, arxiv.org/abs/1808.00320.

¹⁵³ Leah Fessler, „We tested bots like Siri and Alexa to see who would stand up to sexual harassment“, *Quartz*, 22. února 2017, qz.com.

měli k druhým (lidem i ne-lidem) chovat, a dokonce umožní čelit agresivnímu narušování hranic, jež musí každý den snášet ženy a dívky.

Pravidla nic neznamenají, pokud se neprosazují. Toto prosazování působí napětí: jakýkoli pokus omezit chování se nevyhnutelně setkává s odporem kvůli svobodě projevu. V Silicon Valley, aglomeraci, které dal vzniknout národ okouzlený svobodami, stát s individualistickým myšlením a desetiletí bezstarostné alternativní kultury, si cení interní diskuse a svobodu projevu obvykle vnímají jako ryzí dobro. Když Tony Wang ze společnosti Twitter v roce 2012 prohlásil „ve straně za svobodu projevu jsme křídlo prosazující svobodu projevu“, následovalo jen málo negativních reakcí.¹⁵⁴ Dnes jeho prohlášení působí nestydatě politicky, až nezodpovědně. Je-li projev čistě pozitivní, lékem na nevhodný projev je další projev, a sociální sítě jsou tedy údajně řešením vlastních problémů.

Zastánci svobody projevu argumentují tím, že pravda na *trhu myšlenek* přirozeně vyplyne. Je čas se této falešné metafory vzdát. V naší navýsost propojené éře se z tržiště nápadů stala pouliční bouře. Lidé bojují za ideologickou nadřazenost nechutnými prostředky a z vybraných či ukradených informací si vytvoří jakýkoli narativ. Je těžké si představit, jak by se tato situace dala vůbec zvrátit. Prosperující trh vychází z předpokladu, že jsme všichni racionální subjekty, což je optimistický osvícenský názor, který dnes působí nesporně naivně. Ukazuje se totiž, že hluboce zakořeněná ideologie je vůči informacím imunní: v naší rozdělené době podle všeho řada lidí věří jen tomu, čemu věřit chce.

Svoboda projevu se stala ušlechtilým nebezpečím a debata se zvrhla v kulturní válku vedenou podél politických linií obehnaných zákopy. Krajní pravice nedávno vymyslela scénář odporu: podněcujte k nenávisti, pak se ujměte role marginalizovaného bojovníka za svobodu projevu a moderování vykreslete jako cenzuru a fanatismus jako morální křížovou výpravu. Všeobecné opatření, jež chránilo svobodu projevu, se vztahuje jen na vládu, nikoli na průmysl, a má omezenou působnost. První dodatek americké ústavy se samozřejmě týká pouze jedné země a monopolistické

¹⁵⁴ Josh Halliday, „Twitter’s Tony Wang: We are the free speech wing of the free speech party“, *The Guardian*, 22. března 2012, theguardian.com.

tendence moderních technologií (důsledek Metcalfeova zákona¹⁵⁵ a deregulace) znamenají, že pokud je někdo vyloučen z dominantní soukromé platformy, má přiměřený morální nárok tvrdit, že je porušeno jeho právo na svobodu projevu.

Samotná svoboda je složitý koncept. Isaiah Berlin rozlišoval dvě formy svobody: negativní a pozitivní.¹⁵⁶ Negativní svoboda je *svoboda od něčeho*, například svoboda od náboženského pronásledování; pozitivní svoboda je *svoboda k něčemu*, například svoboda vzít si jakoukoli plnoletou dospělou osobu. Michel Foucault šel ještě dále a tvrdil, že svoboda je spíše sloveso než podstatné jméno, spíše praxe než statická vlastnost. Vzhledem k tomu, že toto prosté slovo je zatížené významem, je v etice někdy jednodušší mluvit o štěstí, újmě, povinnosti či ctnosti než o svobodě.

Pokud jde o svobodu projevu, je tu jasný definiční konflikt: (pozitivní) svoboda projevu stojí proti (negativní) svobodě od škodlivých projevů. V technologickém světě nastal čas vychýlit jazýček vah směrem k negativní svobodě. Projev nebyl nikdy zcela svobodný. A ani by neměl být. I ty nejtolerantnější režimy některé projevy definují jako škodlivé, a tedy nezákonné. Obvykle se zakazují věrohodné hrozby násilím, podvodné finanční poradenství či podněcování k nenávisti: těmto škodlivým projevům brání i většina technologických platforem. Dosah a rozsah moderních technologií však všechny druhy napadání obvykle pouze umocňuje. Újma záleží jak na závažnosti nadávek, tak na jejich míře: řadu lidí (zejména ženy a menšiny) ze sociálních sítí vyhnalo nejen množství slovních útoků, ale také věrohodné hrozby. Technologické firmy musí zvážit, zda nebudou slovní útoky zakazovat, i když tyto útoky žádnou věrohodnou hrozbu násilí neobsahují.

Nebezpečí nečinnosti ilustruje *paradox tolerance* Karla Poppera. Pokud nezasáhneme proti netolerantním lidem, dáváme jim moc, jíž mohou bez povšimnutí ničit tolerantní osoby. Přílišná tolerance představuje pro samotnou toleranci ohrožení.

¹⁵⁵ Tzv. Metcalfův zákon, pojmenovaný po jednom z autorů ethernetu Robertu Metcalfovi, říká, že hodnota sítě je přímo úměrná druhé mocnině počtu jejích uživatelů. (Pozn. překl.)

¹⁵⁶ Isaiah Berlin, „Two Concepts of Liberty”, in *Four Essays on Liberty* (Oxford University Press, 1969). (Česky: *Čtyři eseje o svobodě*. Přel. Martin Pokorný. Praha: Prostor, 1999 – pozn. překl.)