

**Univerzita Karlova**

**Filozofická fakulta**

**Ústav informačních studií a knihovnictví**

**Studia nových médií**

**Implementace syntézy hlasu v prostředí vybraných redakcí českých  
médií**

*The Implementation of Voice Synthesis in selected Czech media newsrooms*

Diplomová práce

**Bc. Jiří Labanc**

Praha 2025

Vedoucí práce: Mgr. Josef Šlerka, Ph.D.

Filozofická fakulta univerzity Karlovy  
Studia nových médií

## **Poděkování**

Za podporu děkuji mé rodině a přátelům, které jsem nutil poslouchat můj téměř každodenní pokrok v počtu napsaných znaků. Děkuji rovněž vedoucímu této diplomové práce Josefu Šlerkovi za nadhled a čistě náhodná setkání v obchodech a na tramvajových zastávkách, z kterých se vyklubaly plnohodnotné konzultace. Mé poděkování patří také všem respondentům – Dianě Hlaváčové, Dalii Moudré, Lukáši Wernerovi, Jiřímu Špačkovi, Josefu Vopičkovi a Jiřímu Vrobelovi.

## **Čestné prohlášení**

Prohlašuji, že jsem diplomovou práci vypracoval samostatně, za řádného citování použitých pramenů a literatury. Rovněž prohlašuji, že diplomová práce nebyla použita v rámci jiného vysokoškolského studia či za účelem získání jiného nebo stejného titulu.

V Praze dne 10. května 2025

Jiří Labanc

Filozofická fakulta univerzity Karlovy  
Studia nových médií

## **Klíčová slova**

umělá inteligence, syntéza řeči, text-to-speech, technologie, žurnalistika

## **Keywords**

artificial intelligence, speech synthesis, text-to-speech, technology, journalism

## **Abstrakt**

Tato diplomová práce se zaměřuje na implementaci technologie syntézy hlasu ve vybraných redakcích českých médií. Zkoumá, jak veřejnoprávní i soukromá média s odlišnými modely financování a fungování pracují s nasazením řečových technologií, specificky syntézy hlasu, do svého provozu. Výzkum byl proveden kvalitativní metodou polostrukturovaných rozhovorů s pracovníky a pracovníky výzkumného a produkčního procesu, podílejících se na vývoji technologií umělé inteligence. Následně byla provedena analýza výpovědí respondentů. Výsledky poukazují na vzrůstající zájem o implementaci řečové technologie, jednotlivé redakce se snaží rozšířit dostupnost svého obsahu pro cílové publikum, mimo jiné v závislosti na finančních a technologických možnostech. Z výsledků je patrný vzrůstající důraz na in-house vývoj a syntézu hlasů osob spojených s prostředím zkoumaných mediálních organizací.

## **Abstract**

This thesis focuses on the implementation of voice synthesis technology in selected Czech media outlets. It examines how public service and private media with different funding and operating models work with the deployment of speech technologies, specifically voice synthesis, in their operations. The research was conducted using a qualitative method of semi-structured interviews with research and production staff involved in the development of AI technologies. Subsequently, an analysis of the respondents' statements was conducted. The results point to an increasing interest in the implementation of the technology under consideration, with individual editorial offices seeking to expand the availability of their content to the target audience, depending, among other things, on financial and technological possibilities. The results show an increasing emphasis on in-house development and a synthesis of the voices of those associated with the backgrounds of the media organisations surveyed.

**Rozsah práce:** 150 831 znaků

## Předmluva

Relevance syntetických hlasů roste s vývojem technologií umělé inteligence a digitalizace prostupující do všech oblastí každodenního života. V této diplomové práci jsem se vydal cestou přístupu zástupců mediálních organizací, nikoli samotných novinářů a moderátorů, kterým nové technologie mohou na jedné straně pomoci, na straně druhé nahradit jejich rutinní práci. Přestože se komplexního tématu dotkne i tato diplomová práce, podrobný výzkum těchto i jiných obav a scénářů přenechám jiným badatelům. Zajímá mě uplatňovaná politika médií vůči řečovým nástrojům umělé inteligence, především technologie syntézy řeči, jak tyto nástroje konkrétní média zavádějí a jakou mají redakcích budoucnost. Praktická část diplomové práce bude také zaměřena na syntézu hlasu z pohledu principu jejího fungování a vývoje v kontextu českého jazyka. Protože lze syntetické hlasy vyvíjet řadou způsobů, zaměřím se na standardně využívanou architekturu neuronových sítí. K tomuto účelu využiji expertízy odborníku ze soukromé sféry, kteří jsou úzce spjati s poskytováním a vývojem nástrojů řečových technologií v mediálním odvětví. Plné znění pořizovaných rozhovorů bude k nahlédnutí v mém osobním archivu jakožto autora této diplomové práce.

Kvalitativní výzkum spolu se zvoleným tématem však vedle mnoha pozitiv provází také určitá rizika. Počet respondentů je pro toto konkrétní téma omezen, a z časových či jiných důvodů mně – menšina z nich – na opakované žádosti o rozhovor či poskytnutí expertízy neodpověděli. Čtenáře na tomto místě rovněž upozorňuji, že téma automatizované žurnalistiky zahrnuje mnoho oblastí žurnalistiky. Je nutné brát v potaz, že při mém bádání jsem se zaměřil především na psanou žurnalistiku a využití hlasu, neboť se jedná o domény mně známé.

Během zpracování této diplomové práce také došlo k ukončení pracovního poměru Jiřího Špačka v redakci Seznam Zprávy, proto jsem rozhovor s ním přizpůsobil někdejšímu kontextu jeho působení redakci. Pro aktuálnost a budoucnost řešeného tématu ve společnosti Seznam.cz jsem oslovil Dianu Hlaváčovou.

Vzhledem k povaze studijního oboru Studií nových médií a možností plynoucí ze vzrůstajícího trendu nástrojů umělé inteligence rovněž nemohu ignorovat pozitiva, která tato technologie nabízí. Proto deklaruji, že řečové i generativní AI nástroje jsem využil i při psaní této závěrečné práce.

Konkrétní služby – ChatGPT, Google AI Studio, Elicit, Whisper a UWebASR – mně pomohly na následujících úrovních zpracování textu i zvukových nahrávek:

- Pro vyhledávání studií a rešerší jsem použil nástroj Elicit, díky němuž jsem našel související výzkumy a analýzy i mimo předchozí bádání v databázích vědeckých prací, jako jsou Scopus, ResearchGate a další.
- ChatGPT a Google AI Studio jsem využil k brainstormingu a rozvržení osnovy diplomové práce či shrnutí předložených výzkumů a jejich klíčových částí.
- Nástroje Whisper a UWebASR mně ušetřili hodiny času díky automatizované transkripci pořizovaných rozhovorů.

Zároveň deklaruji, že při jakémkoli užití nástrojů umělé inteligence jsem vždy pečlivě kontroloval a kriticky zhodnotil správnost a úplnost veškerých výstupů.

K cíli za nejlepšími výsledky jsem během psaní diplomové práce využil následující prompty a jejich variace:

- You're a student writing an academic thesis. I will send you the annotation of a thesis entitled The Implementation of Voice Synthesis in selected Czech media newsrooms, which will address the approach of individual newsrooms towards voice synthesis technology in content creation. The thesis includes a theoretical part, in which you have to address topics such as the history of artificial intelligence, automated journalism and its use in newsrooms, ethical issues related to journalists' attitudes towards these technologies, and speech technologies and subfields such as automated transcription, subtitling of TV shows, and especially voice synthesis. The practical part will cover the methodology and approaches of three individual newsrooms towards voice synthesis. Specifically, the newsrooms of Deník N, Czech Radio and Seznam Zprávy. Propose a logical, coherent and substantive structure for such a thesis. The target audience are academic scholars.
- You are a researcher writing a research paper on the use of voice synthesis and speech technologies in journalism. From the following document/text, select the key parts relevant to this topic.
- The following unstructured text that I will send you is a research paper on the use of voice synthesis and speech technologies in journalism. Imagine that you are in the position of a student who has to structure this text into your thesis. The thesis must be of an academic standard and the target audience is academic scholars. When structuring, maintain coherence, correctness of data. Keep all the paragraphs unchanged, only structure the paragraphs logically, coherently, factually and well, and come up with chapter headings.
- You are a critical thinker and new media analyst who focuses on the impact of artificial intelligence on society. You are also a thesis advisor for a student majoring in New Media Studies. Your student

Filozofická fakulta univerzity Karlovy  
Studia nových médií

has sent you the following part of his thesis. Assess the quality of this thesis critically and fairly. Evaluate whether the introduction matches the overall content. Provide your student with substantive and quality feedback on each chapter and on the thesis in general.

## Obsah

Úvod.....	12
Historický vývoj umělé inteligence .....	15
Žurnalistika umělé inteligence.....	17
Implementace automatizované žurnalistiky.....	18
Limity a dopady algoritmické žurnalistiky na novináře .....	22
Řečové technologie v kontextu médií.....	26
Speech-to-text (STT).....	28
Text-to-speech (TTS).....	29
Řečové technologie v českém a slovenském kontextu .....	31
Metodologie .....	35
Stanovení výzkumných otázek.....	35
Kvalitativní výzkum .....	35
Polostrukturovaný rozhovor.....	36
Medailonky respondentů.....	38
Dalia Moudrá (Český rozhlas) .....	38
Diana Hlaváčová (Seznam.cz) .....	38
Lukáš Werner (Deník N).....	38
Jiří Špaček (dříve Seznam Zprávy) .....	38
Josef Vopička (MAMA AI) .....	39
Jiří Vrobel (MAMA AI).....	39
Výzkumná část.....	40
Vývoj syntézy hlasu (MAMA AI).....	40
Shrnutí .....	43
Deník N.....	45
Shrnutí .....	49
Český rozhlas.....	50
Shrnutí .....	53
Seznam.cz .....	54
Shrnutí .....	60
Shrnutí výzkumu.....	61
Závěr .....	66
Zdroje.....	69

Literatura .....	69
Elektronické zdroje .....	69

## Úvod

Velké technologické změny s sebou nesou příležitosti, ale také zasévají zrnko obezřetnosti ve vztahu k vlastní roli uvnitř postupně měnící se společnosti. Ambivalentní projekce budoucnosti doprovázejí současný vývoj umělé inteligence. V celospolečenském pojetí čím dál více skloňovanější technologie je, jak se zdá, změnou přinášející zcela nové paradigma zasahující jak digitální, tak tradiční svět.

Jaké změny nasazení umělé inteligence v mnoha oblastech lidských životů přinese, lze jen stěží, vzhledem k neustálému zvyšování laťky nových schopností, predikovat. Lidská přirozenost má navíc tendence krátkodobé dopady přeceňovat a dlouhodobé podceňovat. Mnohé diskuze mají za to, že technologie zasáhne profese, které podstatnou část stráví za obrazovkami počítače, živí je duševní práce a kreativita. Mezi taková povolání lze jednoznačně zařadit novináře.

Jakkoliv se totiž novináři mohou odkazovat na svou profesní autonomii a kreativní povahu své práce, domnívám se, že žurnalistika je v mnoha případech složena z rutinního schématu pracovního postupu. A rutinu, jak v posledních letech ukazuje vývoj například generativních nástrojů umělé inteligence, lze skutečně automatizovat.

Je to právě obezřetnost postupně přerůstající v obavu z nahrazení lidského charakteru, které mě motivují zkoumat téma automatizované žurnalistiky, neboť představuje příležitost i hrozbu zároveň. Perspektivu automatizace jsem se rozhodl pojmout na příkladu nejkomplexnějšího a také nejpoužívanějšího prostředku lidské komunikace – řeči.

Hlas totiž poskytuje silnější emoce než prostý text. Dovede posluchače zaujmout, pobavit, přesvědčit. S vývojem neuronových sítí a strojového učení se pak zdokonalují nástroje a metody schopné lidský hlas imitovat natolik důvěryhodně, že již nelze s jistotou rozeznat hlas syntetizovaný od toho skutečného. Technologie syntézy hlasu zasahuje též do mediálního odvětví, které si stále více uvědomuje její přednosti. Vývoj zde však ovlivňují mnohdy omezené finanční prostředky i postoje samotných představitelů a členů mediálních organizací. Považuji za důležité zjistit, jakou politiku vybraná média vůči hlasové syntéze v českém kontextu uplatňují.

Vzhledem k neoddělitelnému spojení mezi syntézou řeči, umělou inteligencí a související automatizací žurnalistiky se teoretická část diplomové práce bude věnovat vymezení umělé inteligence a jejímu využití v žurnalistice - tedy algoritmizované žurnalistice. Uvede historický

exkurz umělé inteligence a současný stav automatizované žurnalistiky pohledem mediálních teoretiků.

Syntéza řeči má v žurnalistice široké spektrum využití, od zvýšení dostupnosti obsahu pro zrakově postižené či osoby s poruchami čtení, až po modernizaci způsobu konzumace zpráv pro širokou veřejnost. Média mohou implementovat automatizované předčítání článků na svých webových stránkách a v mobilních aplikacích, čímž rozšíří dosah publikovaného obsahu a otevírají nové možnosti pro personalizaci svého zpravodajství. Nicméně implementace této technologie s sebou nese i etické rozměry, právě včetně obav novinářů o vlastní autonomii a autenticitu žurnalistické práce.

Tato diplomová práce prozkoumá přístup vybraných mediálních organizací vůči technologii syntézy hlasu a jejímu potenciálu automatizovat proces tvorby a distribuci obsahu. Konkrétně využije hlavní výzkumnou otázku: „*Jaký postoj vybrané české redakce uplatňují k využívání syntézy hlasu?*“

Diplomová práce se bude snažit najít odpověď také na dílčí výzkumné otázky: „*Jaké nástroje a služby mediální organizace využívají pro zpracování syntetických hlasů? Jakou roli bude syntéza řeči hrát v budoucím fungování redakcí?*“

Pro účely této práce byl vybrán metodologický přístup v podobě kvalitativních polostrukturovaných rozhovorů s představiteli vybraných mediálních organizací: Deníku N, Českého rozhlasu a technologické společnosti Seznam.cz, která ve své mediální divizi sdružuje redakce řady samostatných médií. Dle časových možností i relevance pro tuto práci jsem vybral konkrétní účastníky výzkumného a produkčního procesu, kteří se na vývoji technologií umělé inteligence v daných redakcích podílejí. Konkrétně jde o projektovou manažerku Českého rozhlasu Dalii Moudrou, zástupce šéfredaktora Deníku N Lukáše Wernera, produktovou manažerku ve společnosti Seznam.cz Dianu Hlaváčovou a bývalého ředitele pro rozvoj v redakci Seznam Zprávy Jiřího Špačka.

Vzhledem ke komplexnosti samotné technologie syntézy řeči jsem se rozhodl využít expertizu odborníků ze společnosti MAMA AI, která se vývojem řečových technologií zabývá a spolupracuje s řadou českých médií – včetně těch, s kterými operuje tato diplomová práce. Josef Vopička a Jiří Vrobel vysvětlí syntézu řeči na fundamentální úrovni a jaké možnosti využití řečové technologie umožňují či jaká rizika přinášejí.



## Historický vývoj umělé inteligence

Ještě před prvotním vyřčením známého pojmu navrhli roku 1943 profesor Warren McCulloch se studentem Walterem Pittsem matematický model neuronu, budoucí základní stavební jednotku neuronové sítě (zvanou SNARC), kterou roku 1951 sestavili Marvin Minsky a Dean Edmons. (Cyberneticzoo, 2009) A byť později sám Minsky v knize *Perceptrons* poukázal na limity a bezcennost neuronových sítí, dnes patří k nejvýznamnějším oblastem umělé inteligence. (Koubský, 2019)

Roku 1956 se na Dartmouthské konferenci sešli elitní výzkumníci z oblastí matematiky, počítačové vědy i kognitivní psychologie, aby mimo jiné zjistili, jak naučit stroje řešit problémy do té doby vyhrazené lidskému myšlení. Na konferenci se etabloval samotný pojem umělá inteligence (artificial intelligence), jehož autorství se přisuzuje počítačovému a kognitivnímu vědci Johnu McCarthy, a tehdy vznikla i samostatná vědní disciplína.

Za rozvoj oboru se zasloužili také Alan Turing, jehož jméno nese i slavný Turingův test, Claude Shannon jako tvůrce vůbec první matematické teorie informace či Ray Solomonoff coby zakladatel teorie algoritmické informace.

Spolupráce významných vědců přinesla optimistické předpovědi – například Herbert Simon roku 1958 prohlásil, že počítač do deseti let porazí člověka v šachu. To se nakonec stalo až o téměř třicet let později. V mezidobí však došlo k útlumu, kterému se zpětně přezdívá zima umělé inteligence. Prvotní optimismus se totiž ukázal jako nadnesený a nadšení vystřídala zklamání. Vývoj neuronových sítí zpomalovala omezení tehdejšího hardwaru, kombinatorická exploze i neúspěch v podobě strojového překladu. Z těchto i řady dalších důvodů došlo k prudkému poklesu financování dalšího AI výzkumu. (Koubský, 2019)

Zájem o AI a neuronové sítě znovu oživilí až výzkumníci, kteří se zaměřili na praktické využití v oblastech počítačů, internetu a komerční sféry, to především díky výkonnějšímu hardwaru a sestavení nových algoritmů. Takzvaný buzzword umělé inteligence nicméně v široké společnosti zůstává již od 50. let, kdy se cílem výzkumu AI stalo vyvinutí AGI - obecné umělé inteligence (z *angl. Artificial general intelligence*), jež bude schopna provádět veškeré úkoly, které dnes zvládnou pouze lidé.

Umělá inteligence zahrnuje podoblasti, jako je strojové učení (*machine learning*), učení pod dohledem (*supervised learning*), učení bez dohledu (*unsupervised learning*), generování přirozeného jazyka (*NLG*) nebo zpracování přirozeného jazyka (*NLP*). Všechny tyto obory, ačkoli jsou označovány jako „AI“, čerpají z různých atributů a vykazují různé schopnosti. Rozdělení na slabou a silnou AI rozlišil už John Searl, který navrhl i myšlenkový experiment zvaný argument čínského pokoje, dokazující, že stroje ve skutečnosti nemohou myslet po vzoru člověka, a silná AI tedy podle Searla neexistuje.

Definice umělé inteligence však dodnes kvůli posunujícím se významu není jasně daná, nejvíce přijímaná je však ta od Marvina Minskeho – jde o vědní obor, jehož úkolem je naučit stroje dělat věci, které vyžadují inteligence, jsou-li prováděny člověkem. (Britannica, 2024)

Oxfordský slovník umělou inteligenci definuje coby „*schopnost počítačů nebo jiných strojů projevat či simulovat inteligentní chování*“ (Oxford English Dictionary, 2025), přičemž slovu inteligence přisuzuje „*schopnost učit se, chápat a logicky o věcech uvažovat.*“ (Oxford English Dictionary, 2025)

V takovém smyslu tedy nelze ani dnešní umělou inteligenci považovat za silnou neboli obecnou, nýbrž coby slabou AI. Jedná se o nové technologie, které umožňují optimalizaci, autonomii a automatizaci a s tím spojené získávání dat a analýzu. S obecnou umělou inteligencí mají pramálo společného, neboť jde o logické automaty pracující na statistickém základě. Široká veřejnost si však na přítomnost termínu za několik desetiletí natolik zvykla, že v povědomí zůstává dodnes a kvůli snadnějšímu pochopení bude obecný pojem umělá inteligence využívat i tato diplomová práce.

## Žurnalistika umělé inteligence

Historický vývoj technologií odjakživa práci novinářů silně determinuje. (Pavlik, 2000) Do popředí diskuze se dnes vzhledem k vývoji vědního oboru umělé inteligence dostávají pojmy jako algoritmizovaná, automatizovaná žurnalistika či žurnalistika umělé inteligence.

Graefe (2016) algoritmizovanou žurnalistiku definuje jako proces využívání softwaru či algoritmů k automatickému generování zpravodajských příspěvků bez zásahu člověka. Kotenidis a Veglis (2021) namítají, že definice takto ohraničená, ač popisuje pravděpodobně nejdůležitější aspekt algoritmické žurnalistiky – tedy automatickou tvorbu obsahu –, opomíjí další aplikace zmíněných technologií v oblasti žurnalistiky. V důsledku tohoto zkrácení je dle nich termín algoritmická žurnalistika obvykle zaměnitelná s celou řadou podobných termínů. Podle Moravce (et. al., 2020) je automatizovaná žurnalistika „*sofistikovanými algoritmickými procesy transformující rozsáhlá data do zpravodajských článků bez zásahu člověka.*“ Zároveň upozorňují na problematické uchopení obecného pojmu umělá inteligence, který, jak bylo zmíněno v předchozí kapitole, je doposud teoretickou hypotézou, mnoho let vzdálenou od praktického konceptu.

Deuze a Beckett (2022) zase rozdělují umělou inteligenci v širším pojetí, kdy jde o zastřešující pojem pro řadu technologií, jako je „*automatická statistická analýza dat, strojového učení a zpracování přirozeného jazyka.*“ Dodávají, že práce novinářů je v mnoha ohledech úzce spjata s rolí AI ve zpravodajství. Interakci mezi lidmi a stroji zdůrazňuje Marconi (2020), podle jehož perspektivy z takovéto spolupráce plynou ve zpravodajských organizacích konkrétní výsledky.

Carlson (2015) zmiňuje, že automatizovaná či algoritmická žurnalistika je průsečíkem mezi mediálním odvětvím a rostoucím důrazem na analýzu dat, známou pod slovním spojením „big data“. „*Na rozdíl od tradiční datové žurnalistiky, která se při analýze a interpretaci dat spoléhá na lidské novináře, algoritmická žurnalistika odstraňuje lidskou účast při psaní a strukturování zpravodajských článků,*“ říká Carlson. Žurnalistika je ostatně důležitá pro samotné datasey, z kterých se AI systémy učí. Mediální obsah je zahrnut v deseti nejvýznamnějších doménách určených pro trénování jazykových modelů společnosti Google. (Schaul et al., 2023)

V širším smyslu tedy zmíněné termíny označují totéž, a nebude-li specifikováno jinak, pro zjednodušení s nimi bude v dané perspektivě pracovat i tato diplomová práce.

## Implementace automatizované žurnalistiky

Minuty i vteřiny dnes rozhodují o tom, komu se podaří informovat o nejnovějších zprávách jako prvním, přičemž ve hře je získání pozornosti konzumenta takového obsahu a z toho plynoucí příjem z reklam. Novinářská profese má proto zavedené postupy tvorby zpravodajského obsahu. Typicky užívanému modelu se říká obrácená pyramida, kdy zpráva odpovídá na základní novinářské otázky a dále doplňuje kontext. Důležitost informací směřuje sestupně. Fakticky jde o rutinní proces, jenž zkušenější novináři provádějí automaticky.

Francesco Marconi, který ve své knize *Newsmakers* (2020) popisuje počátky automatizace a vzestup algoritmičké žurnalistiky v USA, vysvětluje, že novinář vymýšlí témata na články prostřednictvím výzkumu a analýzy dat, shání zdroje a respondenty, s kterými vede rozhovory, následně získané informace shrne a ve spolupráci s editory napíše články. Takovýto postup novinářské práce nazývá lineárním modelem. Lze jej shrnout do třech fází - sběr dat (*newsgathering*), tvorba obsahu (*production*) a jeho distribuce (*distribution*). Co je rutinní, je také strojově automatizovatelné, a umělá inteligence - jak bude popsáno v následujících kapitolách - proces sběru dat, tvorby obsahu i distribuci urychluje. V souvislosti s ústředním tématem knihy tak Marconi předestírá spolupráci člověka a stroje – novinářovy intuice a umělé inteligence –, která proces tvorby obsahu dynamizuje a zefektivní. Marconi tedy mluví o dynamickém modelu fungování redakcí. (2020, s. 27-30)

Jedním z prvních pokusů o algoritmizaci žurnalistiky je případ americké technologické společnosti Narrative Science, která v roce 2010 začala prostřednictvím nástroje Quill pro americkou televizní síť Big Ten Network generovat stručné souhrny zpráv o baseballových a softballových utkáních. Na základě přístupu k datovým souborům se strojové učení na sportovních textech naučilo psát podobné texty, a sportovní výsledky titul zveřejnil v rámci jedné až dvou minut po skončení konkrétního zápasu. Téhož roku navíc Quill našel využití i v sepisování shrnutí a analýz amerického fotbalu a basketballu. Systém dokonce předčil očekávání provozovatelů, a meziročně se návštěvnost webových stránek Big Ten Network zvýšila díky aktuálnosti výsledků o celých čtyřicet procent. V důsledku užití strojových postupů se neprodleně vydané články umísťují na přední pozice indexu internetových vyhledávačů. (Moravec, 2020, s. 81)

Tiskové agentuře Associated Press zase prvky umělé inteligence dvanáctinásobně zvýšily počet publikovaných článků o hospodářských výsledcích amerických společností kótovaných na burze,

a díky nim mohla tisková agentura věnovat pozornost i menším podnikům. (Moravec et. al., 2020, s. 39)

Agentura Reuters pak v roce 2016 představila svůj dva roky vyvíjený nástroj News Tracer, který detekuje nejnovější zprávy na sociální síti X (*dříve Twitter*). News Tracer příspěvky v reálném čase ověřuje, filtruje spam a seskupuje podobné zprávy do jednotlivých „clusterů“ na základě podobných slov. Následně News Tracer tyto clustery klasifikuje do jednotlivých témat a o každém z nich generuje krátké shrnutí pro další zpracování. (Thomson Reuters) Ještě o dva roky dříve začala tisková agentura Associated Press využívat k psaní článků nástroj Wordsmith vytvořený společností Automatic Insights. (Moravec, 2020, s. 82-83)

Moravec (et al., 2020) ve výzkumu z roku 2020 objasňují, že mediální organizace v anglicky mluvících zemích využívají prvky AI k tvorbě obsahu častěji než redakce v České republice. *„Důvodů pro to může být několik. Od omezeného počtu zpravodajských textů, na které by bylo možné prvky strojového učení aplikovat, až po relativně vysoké náklady na vývoj kreativního softwaru, na který v rozpočtech českých zpravodajských médií nezbývají peníze.“* Obdobně hovoří i Cook (et. al., 2021, s. 5), který doplňuje, že hlavní překážkou rozvoje AI v žurnalistice je soutěž o nalákání talentů, ale také udržení stávajících profesionálů v redakcích nabízejících nižší platy ve srovnání s technologickým průmyslem.

První mediální organizací, která v České republice začala pracovat s automatizovaným generováním zpravodajského obsahu, byla Česká tisková kancelář. Roku 2018 takto pokrývala komunální a senátní volby. Moravec (et. al. 2020, s. 41) popisují, že výsledky voleb průběžně zveřejňoval Český statistický úřad a počítačové algoritmy údaje umísťovaly do přednastavené šablony. Takto prošlo kontrolou editorů přes dvě stě automaticky generovaných zpráv. Nasazení předcházely dvouměsíční přípravy, během kterých agentura testovala funkčnost algoritmů na datech z předchozích voleb. Generované zprávy, označené redakční zkratkou „*rur*“, v současné době pokrývají také ceny pohonných hmot a statistiky dopravních nehod.

Moravec (et. al. 2020, s. 42) míní, že automatizované zpracování textů lze rozdělit do dvou oblastí. *„Jednoduché algoritmy na základě šablon využila ve zmíněných případech ČTK. Proces je podobný vyplňování formuláře, a některé podmíněné prvky lze vybírat z konečné množiny vět na základě hodnot údajů. Druhou sférou je strojové učení, podoblast umělé inteligence, která může zahrnovat zpracování jazyka (struktura, gramatika, styl), zpracování přirozeného jazyka (NLP),*

*výběr a způsob prezentace dat nebo porozumění textům,*“ objasňují s tím, že výhodami umělé inteligence jsou snížená chybovost, rozsah zpracovaného obsahu a rychlost.

Láb (2019) uvádí několik oblastí využití automatizované žurnalistiky: sběr dat, monitorování zdrojů a sociálních sítí v reálném čase, datový fact-checking, zpracování či analýza velkých datových souborů, automatizovaná tvorba textu, videa, grafiky, infografiky a cílená distribuce.

Také Kotenidis a Veglis (2021) popisují využití algoritmů v žurnalistice: automatizovanou tvorbu obsahu, vytěžování dat, distribuce zpravodajství a optimalizaci obsahu. Rovněž zmiňují příklad z minulé dekády, nástroj pro automatickou tvorbu obsahu Quakebot amerického deníku Los Angeles Times.

Potenciál technologie umělé inteligence spočívá v odhalování informačních trendů, shromažďování informací, vývoji doporučovacích systémů, ověřování dezinformací a také může rozvíjet investigativní žurnalistiku. Odborníci považují za dvě prioritní oblasti jejího využití právě generování obsahu a jeho následnou personalizaci. (Noain-Sánchez, 2022)

Deuze a Beckett (2022) říkají, že technologie umožňuje pochopit chování a potřeby publika prostřednictvím personalizace obsahu či zkoumání nových možností finančních příjmů. Staronovým polem působnosti užití AI by mohly být sociální sítě, skrze něž podstatná část uživatelů internetu konzumuje zpravodajský obsah. (Kotenidis a Veglis, 2021) O ambivalentním vztahu digitálních médií závislých na sociálních platformách se však hovoří již roky a hořkou pachutí digitálních titulů po neúspěšné spolupráci s těmito platformami glosovalo vícero autorů. (Hrdina, 2023)

Automatické generování zpravodajských textů před více než deseti lety vedlo podle Moravce (et. al., 2020) k novému pojmu „*umělá inteligence v žurnalistice*“. AI novináře postupně vytěsňuje z analýz velkého množství dat a obrazů, formuluje zpravodajské texty, umísťuje a řadí články na zpravodajských serverech. „*Nepotrvá dlouho a bude zřejmé, zda údobí internetové žurnalistiky vystřídá etapu žurnalistiky automatizované či umělé inteligence,*“ píše Moravec. (2020, s. 80)

Další zlom přišel o dva roky později, když společnost Open AI uvedla nástroj generativní umělé inteligence ChatGPT. GPT (*Generative Pretrained Transformer*) je model strojového učení trénovaný na velkých datových sadách lidmi napsaného textu. Svým jednoduchým rozhraním tak firma Open AI demokratizovala generativní AI pro běžného uživatele a otevřela tak

celospolečenskou diskuzi o umělé inteligenci. ChatGPT a jeho obdoby jsou nicméně součástí širšího přechodu ke generativní umělé inteligenci, neboť stejně jako v případě textu, AI se používá také k tvorbě obrazu, videí a zvuku. Zájem o novou technologii dnes plní novinové titulky, na pozadí mediálního odvětví však nutí novináře zamýšlet se nad budoucností své profese. Ve srovnání s jednoduchými sportovními zprávami, založenými na šablonách generovaných v minulé dekádě, je ChatGPT a podobné generativní nástroje výrazným skokem vpřed. Naproti novým schopnostem technologie stojí dlouho vyzdvihované analytické, jazykové a tvůrčí schopnosti coby základní profesní znalosti definující novinářské řemeslo, které dlouho zůstávalo neohroženo dřívější formou strojově psaného zpravodajství. (van Dalen, 2024) Generativní AI své uplatnění nachází rovněž ve shromažďování, produkci a distribuci zpráv – tedy ve zmíněném lineárním modelu fungování profese novináře.

Zmíněná Česká tisková kancelář z důvodu rozšíření generativní AI připravila pro své zaměstnance interní pravidla pro práci s těmito nástroji. Umělá inteligence v redakci funguje na pozici asistenta. Zaměstnancům ČTK doporučuje se s nástroji obezřetně seznamovat. (ČTK, 2023) Další tuzemské tituly následovaly spuštění nástroje ChatGPT vydáním etických kodexů při používání umělé inteligence ve své práci. To poukazuje na serióznost, s jakou novináři k novým technologiím přistupují.

Trend využití umělé inteligence se momentálně pohybuje v hybridní rovině – prvky AI člověka doposud nenahradily, nýbrž dochází k symbiotickému vztahu člověka a počítače. Novináři tedy mohou těžit z výhod, které jim AI nabízí, a rutinní práci přenechat stroji. Nic Newman a Federica Cherubini (2025) v globálním reportu *Journalism and Technology Trends and Predictions 2025* podotýkají, že již v předchozím roce některé větší redakce vytvořily specializované pozice pro práci s AI. Ne jinak je tomu i v případě českých mediálních organizací.

Dosavadní zjištění poukazují na potenciální disruptivní charakter technologie strojového učení v mediálním odvětví. Následující kapitola poukáže na limity žurnalistiky umělé inteligence a prozkoumá výzvy i rozpolcenou ostražitost novinářského prostředí vůči zavádění nástrojů automatizace.

## Limity a dopady algoritmické žurnalistiky na novináře

Nástroje AI mají potenciál v mediálním odvětví generovat články rychleji a levněji než novináři. Užívání umělé inteligence k tvorbě obsahu nicméně vyvolává otázky, zda takovýto obsah může skutečně odpovídat zásadám odpovědné žurnalistiky, a zdali může umělá inteligence nahradit práci novinářů.

V obecné rovině Graefe (2016) popisuje limity algoritmů následovně: *„Automatizovanou žurnalistiku nelze použít v oblastech s chybějícími či nekvalitními daty. Algoritmy navíc získávají poznatky z dat pomocí předem definovaných pravidel a statistických metod, avšak nemohou vysvětlit nové jevy nebo stanovit kauzalitu. To znamená, že algoritmy sice mohou popsat, co se děje, nemohou ale poskytnout výklad, proč se konkrétní věci dějí. Algoritmy tak mají omezenou schopnost pozorovat společnost a plnit žurnalistické úkoly, jako je tvorba veřejného mínění a orientace.“* Graefe rovněž podotýká, že automatizace pravděpodobně nahradí novináře, kteří se věnují pouze rutinním tématům, trend však povede k rozvoji nových pracovních míst.

Od doslovného boomu generativní umělé inteligence na podzim roku 2022 řeší mnoho výzkumů právě její dopad na mediální odvětví, především na případu nástroje ChatGPT. Výzkumná organizace *Center for News, Technology & Innovation* ve své zprávě z kraje roku 2025 uvádí, že používání generativních nástrojů v redakci s sebou nese riziko nepřesností, etických problémů a podkopávání důvěry veřejnosti. Výzkum (Gutiérrez-Caneda, 2023) zabývající se nástrojem ChatGPT poukázal na jeho jazykové bariéry, limitované znalosti, předsudky a netransparentnost.

Experti, akademici a novináři ve studii Noain-Sánchez (2022) popisují dvojí realitu: *„Na jedné straně umělá inteligence přináší mediální společnosti nové tržní příležitosti a zvyšuje novinářovu spokojenost se svou prací. Na stranu druhou nástroje AI zvyšují u novinářů antipatii a strach, v důsledku čehož narůstají i etické problémy kolem AI.“* Oslovení respondenti míní, že příčinami nedůvěry novinářů vůči novým technologiím jsou jejich omezené znalosti, nedostatek dovedností, obava ze ztráty zaměstnání či přesvědčení, že nástroje umělé inteligence naruší podstatu samotné žurnalistiky. Aby se posílila důvěra v technologie, je dle Marconioho (2020, s. 114) zapotřebí novinářům i inovátorům vyvíjející konkrétní nástroje vysvětlovat, na čem a jak druhá strana pracuje.

K obavám novinářů ze ztráty zaměstnání se vyjadřují i Deuze a Beckett (2022). Uznávají, že umělá inteligence dokáže automatizovat mnoho opakujících se a vzorových úkolů, jako je zpracování dat či přepis původních článků: „*Toto vytěsňování umělou inteligencí má již svou historii, která však zůstává neviditelná: Například vyhledávač Google nahradil velkou část namáhavé práce novinářů při rešerši. Může být sporné, že se tolik základního pátrání po faktech nebo zdrojích předává technologii, ale je nepopiratelné, že se zvýšila efektivita, pohodlí a kapacita. Odpůrci strojů zároveň tvrdí, že velká část novinářské práce je přirozeně ‚lidská‘, protože zahrnuje úsudek, kreativitu a dovednosti, které nelze naprogramovat. Tento binární redukcionismus je částečně způsoben omezenou gramotností v oblasti umělé inteligence,*“ tvrdí Deuze a Beckett s tím, že umělá inteligence by neměla být chápána pouze jako vytěsňovací mechanismus, ale spíše jako nástroj pro zvýšení efektivity a pracovní augmentace žurnalistů. Technologie dle autorů v žurnalistice vytváří nové role, například editora algoritmů, který dohlíží na etickou stránku a strojově generovaný obsah.

Gunhild Ring Olsen (2023) v roce 2022 zkoumala postoje členů redakcí v Norsku vůči zavádění automatizované žurnalistiky. Kvalitativní rozhovory s redaktory, editory, projektovými manažery a vývojáři dvou norských redakcí poukázaly na absenci spolupráce mezi vývojáři a novináři, mimo jiné z důvodu, že redaktori, kterých se automatizace může dotýkat, tráví většinu svého času v terénu, a nepřijdou tak s jinými skupinami do kontaktu.

Je nutné upozornit, že výzkum byl proveden ještě před uvedením nástroje ChatGPT. Studie popisuje, že zavedení automatizované žurnalistiky nejspíše vede k poklesu dovedností redaktorů, a ani po několika letech, co jedna z redakcí nástroje automatizace využívala, nebyli redaktori oproštěni od rutinních článků a ani jim nebylo poskytnuto více času na rešerši. Od té doby se však situace na daných pracovištích mohla změnit, a byl by zapotřebí další výzkum.

Šetření provedené Ring Olsen (2023) přesto poukazuje na mnoho detailů, mimo jiné na nefunkční obchodní model digitálních médií. „*Změna ekonomických podmínek na institucionální úrovni je pravděpodobně hlavním důvodem, proč redaktori obou redakcí v posledních letech zaznamenali menší rozmanitost dovedností, více práce a větší podíl rutinní práce. Vytváření algoritmů je zároveň nákladné a časově náročné a strojově psané zprávy zatím nepřinášejí zkoumaným redakcím finanční příjmy. Podle jednoho z editorů byl hlavním důvodem nerentability fakt, že automatizované texty nebyly mezi čtenáři oblíbené. Místo generování příjmů a osvobození reportérů od práce se zdá, že investice do robotické žurnalistiky vedla k rušnějším a méně*

*rozmanitým pracovním dnům reportérů.* “ Tento scénář nastínil již Carlson (2014, s. 423), který zmiňuje, že obava z nahrazení novinářů stroji pramení z dlouhodobého snižování počtu zaměstnanců médií kvůli potížím s generováním příjmu z reklam.

Podle Ring Olsen (2023) automatizace žurnalistiky ve zkoumaných redakcích ovlivňuje smysluplnost práce redaktorů, editorů a vývojářů odlišnými způsoby. Zatímco rozmanitost dovedností, autonomie a prostor pro osobní rozvoj se u redaktorů drasticky snížili, u vývojářů a editorů zůstaly tyto aspekty stejné či se dokonce zvýšily. Podle autorky se zdá, že implementace automatizované žurnalistiky vede ke změnám v organizační kultuře, kde jsou technologické nástroje čím dál víc upřednostňovány. Stanoviska jednotlivých skupin však nelze jasně generalizovat. Zatímco některé vývojáře a seniorní editory ovlivňovala podnikatelská logika, kdy redakční kulturu vnímali jako překážku technologického pokroku, jiní respondenti – reportéři, juniorní i seniorní editoři a někteří vývojáři – zastávali tradiční novinářskou logiku, orientovanou na povolání novináře, a zdůrazňovali mnoho dovedností s profesí spojené. Kvůli automatizaci taky novináři pracují ve stále nejistějších podmínkách, implementace žurnalistiky umělé inteligence dle Ring Olsen (2023) zřejmě zpochybňuje přesvědčení a smysl práce novinářů a vyvolává v nich pocit nedostatečnosti, nedoceníení a frustrace.

Ve světě všudypřítomných algoritmů je umělá inteligence tématem pro všechny specializované novináře, a budou muset pochopit, jak se stává součástí jejich oboru. Deuze s Beckettem (2022) však upozorňují, že gramotnost umělé inteligence v žurnalistice není jen o pochopení, jak technologie funguje. Je nutné brát v potaz i rozpoznání důsledků, potenciálu a omezení, které přináší. Mimo jiné technologie hraje natolik systémovou roli, že ji nelze ignorovat ani na úrovni reorganizaci redakce a náborem nových zaměstnanců. Stále je však podle nich důležité rozptýlit přetrvávající představu o AI coby všemocné technologii. Autoři tvrdí, že *„futuristické a spekulativní představy umělé inteligence, inspirované technologickým průmyslem, totiž podbarvují a utvářejí technologický rámeček prezentace samotné technologie. V důsledku toho vyvstávají obavy, zda a kdy se lidé stanou zastaralými a zároveň přijmou nevyhnutelné scénáře, tak jak si je představuje relativně neregulovaný technologický průmysl.“*

Bez gramotnosti v oblasti umělé inteligence novináři riskují, že přenechají kontrolu nad rolí umělé inteligence v žurnalistice technologickým společnostem a tvůrcům politik, a potenciálně se tak sníží nezávislost žurnalistiky a schopnost sloužit veřejnému zájmu. (Deuze a Beckett, 2022)

Důvěru v nové nástroje podle Marconioho (2020, s. 60) může narušovat i outsourcovaný vývoj technologie mimo mediální dům – například do korporací jako Google či Microsoft –, a redakce by implementaci měly řádně vyhodnotit. Globálně se technologie AI dostávají do mediálních institucí stále více. (Beckett a Yaseen, 2023) Redakce si také mohou vytvářet vlastní AI nástroje a systémy. (Deuze a Beckett, 2022)

Umělá inteligence je dle Deuze a Becketta (2022) nápomocna při řešení jednoduchých rutinních úkolů ve velkém měřítku. V rámci složitých úkolů však trpí celou řadou etických problémů. *„Systémy umělé inteligence jsou náchylné k manipulaci a chybám, mají tendenci být systémově zaujaté a prohlubují a zesilují existující nerovnosti.“*

Carlson (2014, s. 423) podotýká, že i když umělá inteligence umožní novinářům věnovat se smysluplnější práci, neexistuje žádná záruka, že přesunutí novináři najdou alternativní zaměstnání. Zastánci automatizované žurnalistiky nicméně tvrdí, že AI bude prodlouženou rukou novináře, která mu umožní soustředit se na kreativní nebo investigativní práci, mezitím co stroje budou dělat rutinní zpravodajské úkony.

Do diskurzu generativní AI zasahuje také rovina autorství. Vystává otázka, komu ze spolupráce člověka a stroje přisoudit roli autora mediálního obsahu. I pokročilý nástroj umělé inteligence stále bude pracovat pod lidským dohledem, byť bude vykonávat zásadní činnosti. (Noain-Sánchez, 2022) Tal Montal a Zvi Reich (2017, s. 831) podotýkají, že *„podstatou algoritmického procesu robotické žurnalistiky je automatizace vyprávění příběhů, přizpůsobená lidskému chápání života jako sledu souvislých vyprávění.“* Jinými slovy, na tom, zdali bude možné rozpoznat lidského autora od stroje, bude záviset budoucí schopnost stroje „vyprávět“ (z angl. *storytelling*).

Aby mediální průmysl využil potenciálu umělé inteligence, je nutné pochopit a řešit výzvy s ní spojené. To mimo jiné zahrnuje vypracování etického rámce, investice do vzdělávání novinářů o oblasti umělé inteligence a tvorbu předpisů a strategií, díky kterým média dokáží držet krok s rychlým vývojem AI technologií. (Febri Sonni et. al., 2024)

Vývoj redakčních generativních nástrojů může probíhat za účelem tvorby řady obsahů – textu, obrazu, videa či zvuku. Právě poslední zmíněná doména, na níž se tato práce zaměřuje především, bude coby oblast řečových technologií představena v následující kapitole.

## Řečové technologie v kontextu médií

Novináři a mediální organizace využívají řečové technologie již roky. Ruční přepis rozhovorů pomáhají zefektivnit volně dostupné transkripty technologie speech-to-text (STT), titulkování pořadů v televizním vysílání – částečně – zajišťují automatizované systémy. Zpravodajské weby začínají využívat systémy text-to-speech (TTS), které konvertují psaný text do mluvené podoby, a za využití syntetických hlasů redakce nabízejí konformnější konzumaci svého obsahu. Vývoj těchto technologií navíc urychlují i neuronové sítě, které díky strojovému učení tvoří umělý hlas, těžko rozeznatelný od skutečného lidského. Za zaváděním řečových technologií stojí také snaha poskytnout dostupnější zpravodajství pro sluchově či jinak znevýhodněné osoby.

Nic Newman (2018) se v reportu z roku 2018 věnuje roli hlasu v mediálním průmyslu. Přestože se v někdejší kontextu věnuje hlavně zařízením a hlasovým asistentům, jako jsou produkty Amazon Alexa a Echo, Google Assistant či Home nebo Apple Siri, podotýká, že navzdory jejich možnostem pro běžné uživatele tyto hlasové reproduktory skýtají hlubší technologické změny. Již dnes jsou ostatně součástí mnoha zařízení – mobilních telefonů, automobilů, televizorů, ledniček i obecně chytrých domácností.

*„Hlas budeme stále častěji používat k ovládání zařízení a přístupu k médiím, protože je pro mnoho účelů rychlejším a pohodlnějším vstupem než dotykové obrazovky a dálkové ovladače,“* tvrdí Newman (2018). Výsledky průzkumu (Newman, 2018) z Velké Británie odhalují, jak lidé chytré reproduktory využívají, a nejvíce oceňované funkce. Na prvním místě je možnost přehrávání hudby – 84 % respondentů uvedlo, že tuto funkci používají, přičemž 61 % dotazovaných uvedlo, že jde podle nich o nejvíce oceňovanou funkci. Ke zpravodajství lidé využívají speakery v 46 % případů, ale pouze 1 % uvedlo, že se pro ně jedná o nejdůležitější funkci. Konzumace zpráv je dle reportu na nižší úrovni než předpověď počasí, kterou využívá 58 % lidí a 4 % lidí si funkce nejvíce cení.

Spolu s trendy hlasových asistentů, kteří se v minulé dekádě etablovali v mnoha domácnostech, se mediální organizace podle Newmana (2018) zaměřily na budování *„sociálních a rodinných zážitků.“* Jmenuje společnost BBC, která roku 2018 spustila interaktivní dětskou aplikaci umožňující komunikovat s oblíbenými postavičkami z pořadů televizního kanálu CBeebies, provozovaného BBC. Zpráva také zkoumá strategie a přijímané politiky vydavatelů, jako jsou The Wall Street Journal, The New York Times či Sky News vůči technologiím hlasových asistentů.

Podle průzkumu Reuters Institute for the Study of Journalism agentury Reuters z roku 2019 až 78 % respondentů věří, že řečové technologie, které se zabývají problematikou automatického rozpoznávání řeči, počítačovou syntézou či hlasovými dialogovými systémy, v příštích několika letech změní způsob, jakým diváci přistupují k médiím. (Newman, 2019, s. 7)

Marconi a Siegman (2017) nastiňují možný vývoj, kdy řečové technologie budou schopny analyzovat sentiment a emoce v reálném čase. „*Novinář zavolá do PR firmy zastupující majitele továrny, aby si vyslechl jejich verzi příběhu. Má podezření, že jejich zástupce něco tají: technologie analýzy hlasu prohlásí, že tón osoby v telefonu je ‚nejistý‘ a ‚nervózní‘,*“ představují si autoři budoucí práci novinářů. Řečové technologie podle nich pomohou odhalit pocity a emoce respondentů, aniž by je vyřkli.

K předpovědím Marconiho a Siegmána z roku 2017 však vede náročná cesta. Černocký (2024) upozorňuje, že na rozpoznávání emocí si současné technologie stále „*vylamují zuby*“, přičemž k prolomení tohoto nedostatku bude zapotřebí ještě dlouholetý vývoj. Navzdory závěrům některých výzkumů je totiž dle Černockého problémová testovací část takovýchto systémů.

„*Různé evaluace a články sice tvrdí, že nástroje pro rozpoznávání emocí fungují, většinou se však testují na hraných emocích, kdy se pozve skupina herců, aby odehráli smutek, neutrální stav či depresi. Pak vše funguje. Jenomže po nasazení těchto nástrojů do call center jsou hlášeny zásadní nedostatky,*“ tvrdí Černocký (2024) s tím, že limity jsou dány dostupnými daty. Komplexní trénovací data určená pro vývoj modelů na rozpoznávání emocí totiž podle něj téměř neexistují. „*A když už ano a jsou ručně označována, ani lidští anotátoři se neshodnou, zdali určitá emoce znamená smutek či neutrální stav. Jestliže neznáte osobu naproti vám ani kontext, správnou emoci možná vůbec nepoznáte,*“ dodává Černocký. (2024) Ačkoliv je řeč nejefektivnější způsob komunikace, přezkoumat její zvukový signál není snadné. Avšak jak poukáže praktická část této diplomové práce, ani nemožné, a podle některých odborníků lze do jisté míry posuzovat emoce za pomoci počítačových systémů.

Důležitými technologiemi v žurnalistice jsou v kontextu řečových technologií převody řeči na text a z textu na řeč – speech-to-text (STT) a text-to-speech (TTS). Jak obě technologie fungují, bude popsáno v následujících podkapitolách.

## Speech-to-text (STT)

Technologie speech-to-text (STT) převádí mluvená slova na text. V kontextu žurnalistiky jde o významný nástroj, který novinářům šetří hodiny práce s prepisem rozhovorů, diskuzí, projevů a dalších zvukových záznamů. (Marconi, 2020, s. 90)

Marconi (2020) popisuje ještě donedávna trvající realitu novinářů. Jejich práce zahrnovala nahrávání rozhovorů na diktafony, přičemž zpracování probíhalo nahráním souborů do počítače, kde je novinář přehrával, přetáčel a pozastavoval, aby našel konkrétní citaci či pořízený rozhovor přepsal. S rozvojem transkriptorů a dalších AI nástrojů je však možné tuto rutinní a často zdlouhavou práci násobně urychlit tak, že novinář audiosoubor nahraje do systému, který nahrávku přepíše. „*Systémy převodu řeči na text mohou reportérům a producentům pomoci automatizovat časově náročné rutinní úkony, jako je přepis a tvorba titulků. Místo toho, aby reportéři seděli u hodin videozáznamů nebo zvukových rozhovorů a ručně zapisovali, co dotyčný řekl, mohou nyní pomocí umělé inteligence automaticky získat textový přepis,*“ objasňuje Marconi. (2020, s. 90-91)

Mezi využívaný speech-to-text software patří kupříkladu Otter.ai, Colibri.ai, SpeechText.ai či Whisper od OpenAI. Známost českou společností nabízející transkripci a titulkování je firma Beey.io.

Na trh s AI technologiemi vstoupila společnost Plaud AI, která v srpnu roku 2023 představila své zařízení Plaud NOTE a následujícího roku také Plaud NotePin. Tyto kapesní nástroje uživatelé spárují se svým mobilním telefonem a mohou díky nim zaznamenávat zvuk jako na diktafon. Integrovaný Whisper umožňuje přepis záznamů v reálném čase již během schůzky či telefonního hovoru, a zakomponované jazykové modely GPT-4o či Claude 3.5 Sonnet nabízejí možnost s přepsaným textem dále operovat. (Phelan, 2024)

Příkladem užití speech-to-text je i titulkování. Britská veřejnoprávní BBC je do svého vysílání poprvé implementovala už roku 1979. Dnes skrze svou streamovací službu BBC Sounds produkuje až 27 tisíc hodin obsahu měsíčně, a komplexní ruční titulkování z časových i finančních důvodů není možné. (Carter Ross, 2024) Minulý rok proto začala mediální organizace testovat zmíněný Whisper, aby své podcasty a další audioobsah zpřístupnila neslyšícím a nedoslýchavým divákům. Interní diskuze však odhalily také vysokou poptávku po této funkci ze strany zájemců o studium cizích jazyků. Díky otevřenosti tohoto softwaru může BBC do Whisperu implementovat vlastní funkce, jako je diarizace mluvčích (*při níž se určuje, který z řečníků v daný moment promlouval,*

*pozn. aut.*): BBC však stále pracuje s nástroji umělé inteligence velmi obezřetně. Produktový tým BBC již ale dokáže určit nízkorizikový obsah, který umělá inteligence může přepsat bez nutnosti ručních kontrol. (Magee, 2024)

Koenecke (et. al., 2024) sice vyzdvihují velké množství správných přepisů u nástroje Whisper, dodávají však, že 1 % zvuků obsahovalo halucinované fráze a věty, které se ve vstupním zvuku nenacházely. Kolektiv autorů však pracoval s angličtinou, a je patrné, že v jiných jazycích je chybovost (*Word Error Rate, WER*) vyšší. Data ukazují, že v případě Whisperu se – v závislosti na použitých datasetech a modelech – chybovost u angličtiny (při použití modelu large-v3) pohybuje mezi 4,1-9,3 %. *Word Error Rate* se v rámci českého jazyka pohybuje v rozmezí 9-10,1 %. (Jong Wook, 2023) Transkripty dokáží šetřit čas, jejich omezení spojení s přesností automatizovaného přepisu ale stále vyžadují lidský dohled.

### **Text-to-speech (TTS)**

Zatímco speech-to-text často nachází interní využití v rámci redakce, druhá technologie, text-to-speech (*TTS*), v online žurnalistice umožňuje publiku přijímat obsah ve formě audio článků a podporuje moderní trend personalizace obsahu. Díky syntetickým hlasům si uživatelé – potýkající se se zrakovým postižením či poruchami čtení – mohou obsah poslechnout.

Deník The Washington Post využil roku 2020 svůj nástroj Heliograf, aby v rámci někdejších prezidentských voleb za pomoci umělé inteligence zveřejňoval audio výsledky. Funkci deník implementoval do svých vybraných podcastů, kde posluchačům předčítal aktuální informace automatizovaný hlasový asistent. (WashPostPR, 2020) Zvýšený zájem o zvukovou stránku, kterou podchytil, se serveru vyplatila. Čtenáři, jimž The Washington Post v rámci svých mobilních aplikací poslech audiočlánků nabídl, trávili třikrát více času konzumací obsahu The Washington Post. V návaznosti na to pak server integroval i službu Amazon Polly, která zpracovává technologii text-to-speech. (WashPostPR, 2021)

Jedním z nástrojů text-to-speech technologie je také ALTO, který vyvinula britská veřejnoprávní společnost BBC, a umožňuje převádět videoobsah do více jazyků. Celý software může kontrolovat jediný editor. (BBC, 2015) Podobnou technologii využívají i The New York Times, ABC News a ESPN v rámci nástroje Trint. (Marconi, 2020, s. 91)

Příkladem zavádění hlasových technologií je také indický startup Gram Vaani, který vyvinul systém využívající zpracování přirozeného jazyka (*NLP*) k převodu hlasových zpráv v místních jazycích na zpravodajské aktuality. Projekt se tak zapojil do řešení komplexních problémů rozvojových zemí, které se kvůli často omezeným finančním zdrojům potýkají s obtížným zaváděním technologií do novinářského odvětví. (Febri Sonni et. al., 2024)

Zpravodajství veřejnoprávní televize BBC Global News spustilo na webu BBC.com nástroj pro převod textu na řeč s využitím umělé inteligence, vyvinutý ve spolupráci se společností Microsoft. Funkce využívá hluboké neuronové sítě k vytváření přirozeně znějícího hlasu a poprvé se využil v seriálu BBC Worklife „*The Life Project*“. Nástroj se učí uživatelské preference, čímž zajišťuje personalizovaný obsah při zachování redakčních standardů BBC, a je navržen s regionálním přízvukem, který se automaticky aktualizuje při změnách textu. (BBC(b), 2020)

V důsledku zavedení syntézy hlasu do žurnalistiky si diváci vytvářejí hlubší vztah nejen vůči samotnému médiu, ale také vůči konkrétnímu hlasu. Technologie text-to-speech se u spontánní řeči rovněž potýká s nepřesnostmi při zdůrazňování správných slov a problémy s harmonizací řeči ve vztahu k celkovému kontextu. (Reddy in Tomková, 2024) Také může obtížně vyslovovat některá cizí jména. Marconi k tomuto riziku dodává, že „*jazyk a řeč v hlasových zařízeních jsou naprogramovány prostřednictvím konečného procesu fonetizace a prozodie, omezující jejich schopnost porozumět, jak by se jméno mělo vyslovit, pokud neodpovídá parametrům zvoleného jazyka.*“ (Marconi, 2020, s. 90) Některé systémy tak vyžadují dodatečnou pomoc při učení se výslovnosti. Blíže tento problém a jeho řešení popíše praktická část diplomové práce, konkrétně na příkladu českého zpravodajského serveru Deníku N.

Automatizované čtení zpráv využijí hlavně uživatelé, kteří potřebují či preferují zprávy poslouchat než je číst. Vývoj neuronových sítí umožňuje syntéze hlasu natolik významný technologický pokrok, že technologie má již dnes potenciál doplňovat běžné čtení lidských zpráv v rádiu, podcastech i v televizi jakožto voiceover. Podle Cambre (et. al., 2020) se perspektiva TTS ve zpravodajských organizacích bude odrážet na základě „*příjemné zkušenosti se syntetickým hlasem.*“ Jak dodává Heiselberg (et. al., 2022), koncept aplikací předčítající zprávy namísto skutečných moderátorů může také přinést problémy s důvěryhodností, pokud by publikum vnímalo hlas jako nepřirozený nebo odcizující. Studie, která na dánském publiku zkoumala prožívání a důvěryhodnost při čtení zpráv syntetickým hlasem, rozdělila posluchače na dvě skupiny: ti, kteří si uvědomili či tušili, že hlas je syntetický ještě před odhalením tohoto faktu, a posluchači, kteří

syntetický hlas nepoznali. Pro publikum je v první řadě podstatné, aby samotná zpráva a zdroj za vysíláním zpráv – tedy zpravodajská instituce – byli důvěryhodné. „*Je nanejvýš důležité, aby byla upřednostněna transparentnost, neboť je předpokladem důvěryhodnosti: účastníci jednoduše potřebují vědět, zda jde o lidský nebo automatický hlas, který čte zprávy.*” Někteří účastníci šetření však měli pocit, že důvěryhodnost je zpochybněna, jestliže zprávy předčítá umělý hlas. (Heiselberg et. al., 2022)

## **Řečové technologie v českém a slovenském kontextu**

Mezi významné milníky českých médií v oblasti řečových technologií patří projekt Českého rozhlasu Gott navždy. Koordinace iniciativy pro znovuoživení syntetizovaného zpěvákovy hlasu probíhala na dvou úrovních – technologické a eticko-právní.

Digitální otisk hlasu zesnulého zpěváka Karla Gotta vytvořili odborníci ze společnosti SpeechTech ve spolupráci s Výzkumným centrem NTIS. Pro syntézu Gottova hlasu měli k dispozici přibližně dvě stě hodin nahrávek z archivu Českého rozhlasu, avšak kvůli rušivým faktorům (*šumu či hudbě na pozadí, pozn. aut.*) použili z celkového množství desetinu zvukových nahrávek Gottova hlasu. Jakmile výzkumníci po procesu, kdy zvukové záznamy převedli na text a následně párovali „*jednotlivá písmena nebo skupiny písmen v textu a ve zvuku,*“ vytvořili hlasový model. (Český rozhlas, 2023)

Český rozhlas vypracoval striktní interní pravidla, podle nichž hlasovou syntézu využije „*výhradně pro texty a repliky, které dotyčný/dotyčná sám napsal/a nebo pronesl/a.*“ Podle náležitostí lze dílo s hlasovou syntézou použít jedině „*v náležitém kontextu, které odpovídá původnímu účelu díla. Nesmí být vytrženo z kontextu, resp. vloženo do zavádějících souvislostí.*“ (Český rozhlas, 2024)

Hlas, který následně předčítal vybrané pasáže z autobiografie Karla Gotta *Má cesta za štěstím*, namluvil přibližně 16 tisíc slov. Jednalo se navíc o první český pokus využití umělého hlasu v literárně-dramatickém díle. Projekt Gott navždy trval napříč létem 2023, odvysílal jej Český rozhlas Dvojka a stal se nejposlouchanější četbou na pokračování v historii Českého rozhlasu. (Bumbálková, 2023)

Postoj české společnosti vůči technologiím umělé inteligence v žurnalistice pak zkoumala Katedra žurnalistiky Fakulty sociálních věd Univerzity Karlovy ve spolupráci s agenturou IPSOS a Technologickou agenturou ČR. V první vlně sběru dat (říjen 2023) zaregistrovalo projekt Gott

navždy 27 % oslovených. Ve druhé vlně (sesbírané v dubnu 2024) se počet lidí, kteří projekt zaregistrovali, snížil na 24 %. Opačný trend zaznamenalo povědomí o možnosti poslechu audiočlánků na serveru Seznam Zprávy, kde od jara 2023 předčítala články robotka Jana – syntetický hlas poskytnutý českým startupem MAMA AI. V první vlně zaznamenalo robotku Janu 12 % lidí. V následující vlně počet oslovených stoupl na 16 %. Robotickou moderátorku Hacsiko na Rádiu Expres zaznamenalo 7 % lidí (sběr dat o tomto projektu proběhl pouze v druhé vlně). Robotický čtenář článků na webu Hospodářských novin se dostal do povědomí 8 % lidí v obou vlnách. (Moravec, 2024)

Společnost SpeechTech se podílela na projektu ELJABR (*Eliminace jazykových bariér handicapovaných diváků České televize*), jehož cílem bylo zpřístupnit televizní vysílání osobám se sluchovým postižením. Projekt byl zahájen v roce 2006 ve spolupráci s Katedrou kybernetiky Fakulty aplikovaných věd Západočeské univerzity v Plzni a financován Ministerstvem školství, mládeže a tělovýchovy spolu s Českou televizí. SpeechTech se jako spoluřešitel projektu podílel na vývoji a implementaci systémů automatického titulkování, které umožnily generování titulků k živě vysílaným pořadům pomocí technologie automatického rozpoznávání řeči. Stejná technologie může zajišťovat také titulkování offline pořadů. (Müller, 2014)

Během živých relací České televize, například diskuzí či komentování sportovních utkání, je však zvuková stopa obvykle hůře rozpoznatelná kvůli emotivním dialogům řečníků, při kterých si nezdědka skáčou do řeči, mluví nespisovně a spontánně. V takových případech se v ČT využívají takzvaní stínoví řečníci.

*„Originální zvuková stopa pořadu je poslouchána a přemlouvána ve studiovém či kancelářském prostředí tzv. stínovým řečníkem, který původní záznam přemlouvá „neutrálním“ způsobem do mikrofonu systému tzv. automatického diktátu, což je systém automatického rozpoznávání mluvené řeči zapisující v reálném čase diktovaný text. Způsob titulkování s využitím stínového řečníka se používá u živého pořadu Otázky Václava Moravce,“* objasňuje Müller (2014). V současné době SpeechTech zajišťuje titulky k veškerým živě vysílaným pořadům, mezi které vedle zmíněných Otázek Václava Moravce patří Hyde Park, Ekonomika ČT 24, Interview ČT 24, Star Dance, TýTý, Thálie, Sama doma, hokej, fotbal, tenis, biatlon a další. (SpeechTech)

Využití technologií TTS a STT v online světových a slovenských médiích zkoumala Šimona Tomková (2025). *„Technologie speech-to-text může mít problémy se spontánním projevem, který*

*je běžný v novinářských rozhovorech a diskusích. Přepisuje výplňková slova a nepracuje s kontextem výpovědi. Právě z tohoto důvodu je nutné, aby novinář i takto přepsaný text znovu upravil do čtenáři srozumitelného projevu,*“ píše Tomková (2025) s tím, že dostupné nástroje dokáží identifikovat dialogy a osoby v pořízených nahrávkách a také novinářům umožňují rychle vyhledat konkrétní část výpovědi.

Z údajů za září a říjen 2024 vyplývá, že syntéza hlasu navzdory svému potenciálu není ve světových redakcích rozšířená. Tomková (2025) zjišťovala, které z 20 nejnavštěvovanějších světových a slovenských zpravodajských serverů nabízejí publiku možnost čtení článků pomocí neuronových hlasů.

*„Mezi analyzovanými zahraničními médii vyniká BBC, která u některých článků nabízí funkci převodu textu na řeč (TTS) přímo na svých webových stránkách a kromě toho poskytuje robustnější služby prostřednictvím aplikace BBC Sounds. Naproti tomu CNN, The New York Times a The Guardian tuto funkci nenabízejí přímo na svých webových stránkách, ale poskytují ji prostřednictvím specializovaných aplikací, jako jsou CNN Audio, NYT Audio a další.“* (Tomková, 2025)

Naproti tomu jiná významná online média, jako jsou MSN, Fox News, Daily Mail a Forbes žádnou formu předčítání článků umělými hlasy nenabízejí.

Ještě striktnější výsledky poskytla tabulka dvaceti nejvýznamnějších slovenských online médií, z nichž nabízí možnost poslechu článků syntetizovaným hlasem pouze Denník N. Zpravodajský server funkci zavedl v dubnu 2023 a postupně ji rozšířil na všechny články. Dnes již syntetické hlasy předčítají i krátké zprávy v aplikaci denního zpravodajství Denníku N Minuta po minutě. *„Tato funkce je navíc podporována ve speciálním audio přehrávači v aplikaci Denník N, čímž zvyšuje dostupnost pro mobilní uživatele. Zbývající slovenské mediální servery, včetně významných webů, jako jsou Aktuality.sk, Sme.sk, Pravda.sk a další, však na svých webech ani prostřednictvím externích aplikací žádné možnosti hlasového čtení poháněného umělou inteligencí nenabízejí,*“ uvádí Tomková (2025) s tím, že Denník N využívá technologii Azure od společnosti Microsoft (v době dokončení této diplomové práce však Denník N již pracoval na implementaci služby ElevenLabs, která bude blíže popsána v praktické části, v kapitole Deník N, pozn. aut.) Občasné chyby, kterých se syntetické hlasy dopouští, jsou pochopitelné vzhledem k velikosti zdrojových dat pro trénování hlasového modelu ve slovenštině. Jak nicméně popíše praktická část této

diplomové práce, se svou expertizou na pole vývoje českých a slovenských hlasových modelů vstoupil například český startup MAMA AI a další.

Měřítka tohoto výzkumu dle Tomkové (2025) nelze zobecňovat na situaci mezinárodního mediálního trhu, poukazuje spíše na trendy v mediálním odvětví. Výsledky slovenského mediálního trhu však poskytují reprezentativnější náhled, výčet zpravodajských webů totiž zahrnuje v podstatě všechna tamní významná online média, a úspěšná praxe Denníku N ukazuje, jak může zavádění technologií syntézy hlasu probíhat.

Následuje praktická část diplomové práce. Kapitoly budou zaměřeny na český kontext mediální krajiny, konkrétně na vybraná česká média operující s řečovými technologiemi a konkrétně text-to-speech, a na expertízu odborníků k fungování technologie.

## Metodologie

### Stanovení výzkumných otázek

Cílem výzkumu je zjistit, jaký přístup uplatňují vybrané české redakce vůči řečovým technologiím a konkrétně technologii řečové syntézy. Technologie může automatizovat zaběhlé redakční procesy a rutiny, přičemž jak poukázaly předchozí kapitoly, automatizace žurnalistiky vzbuzuje u novinářů nedůvěru i obavu ze ztráty své autonomie a v důsledku i zaměstnání. V následující praktické části tedy bude brán zřetel i na budoucí vývoj syntézy hlasu v redakcích.

Diplomová práce se bude snažit odpovědět na hlavní výzkumnou otázku: „*Jaký postoj vybrané české redakce uplatňují k využívání syntézy hlasu?*“ Obecně položená otázka umožňuje se zaměřit na jednotlivé individuality každé z redakcí a jejich možnosti vývoje.

Výzkum se bude snažit najít odpovědi také na dílčí výzkumné otázky: „*Jaké nástroje a služby mediální organizace využívají pro zpracování syntetických hlasů? Jakou roli bude syntéza řeči hrát v budoucím fungování redakcí?*“

Využití syntézy hlasu i dalších nástrojů umělé inteligence se v jednotlivých redakcích liší. Součástí kvalitativního výzkumu budou média jak veřejnoprávní, tak soukromá.

### Kvalitativní výzkum

Pro dosažení stanovených cílů je důležité přistupovat ke každému zkoumanému subjektu individuálně. Součástí šetření jsou totiž mediální organizace s různými možnostmi vývoje a financování, a pro komplexní popis motivací a postupů autor zvolil kvalitativní výzkumný přístup.

Hendl podotýká, že kvalitativní výzkum nelze jednoznačně vymezit, neboť zahrnuje širokou škálu přístupů a metodologií. Glaser a Corbinová (in Hendl, 2005, s. 49-50) jej považují za výzkum, jehož výsledky nejsou získávány pomocí statistických metod či kvantifikace. Hendl (2005) doplňuje, že jiní autoři s touto definicí nesouhlasí a upozorňují, že jedinečnost kvalitativního výzkumu nespočívá pouze v absenci čísel. Creswell (in Hendl, 2005, s. 50) popsal kvalitativní přístup následovně: „*Kvalitativní výzkum je proces hledání porozumění založený na různých metodologických tradicích zkoumání daného sociálního nebo lidského problému. Výzkumník vytváří komplexní, holistický obraz, analyzuje různé typy textů, informuje o názorech účastníků výzkumu a provádí zkoumání v přirozených podmínkách.*“

Dle Hendla (2005, s. 50) je práce výzkumníka přirovnává k činnosti detektiva. „*Výzkumník vyhledává a analyzuje jakékoliv informace, které přispívají k osvětlení výzkumných otázek, provádí deduktivní a induktivní závěry. Seznamuje se s novými lidmi a pracuje přímo v terénu, kde se něco děje.*“

Hendl (2005, s. 52) dodává, že získaná data se induktivně analyzují a interpretují. „*Kvalitativní výzkumník nesestavuje ze získaných dat skládku, jejíž konečný tvar zná, spíše konstruuje obraz, který získává kontury v průběhu sběru a poznávání jeho částí.*“ Odkazuje na emergentní, tedy pružný charakter kvalitativního výzkumu, díky němuž může výzkumník modifikovat či rozšířit výzkumné otázky v průběhu sběru a analýzy dat.

### **Polostrukturovaný rozhovor**

Diplomová práce pro získání odpovědí na výzkumné otázky a cíle práce využije kvalitativní výzkumný přístup, konkrétně metodu polostrukturovaných rozhovorů. Jak zmiňuje Švaříček a Šed'ová (2007, s. 159), jde o nejpoužívanější metodu sběru dat v kvalitativním výzkumu a jeho cílem je získat detailní a komplexní informace o studovaném jevu. Také apeluje na podstatnost dovedností k vedení rozhovoru, neboť během práce je tazatel nucen dotazovat se na mnoho otázek, a to buď formálně či neformálně. „*Celý proces získávání dat prostřednictvím této metody sestává z výběru metody, přípravy rozhovoru, průběhu vlastního dotazování, přepisování rozhovoru, reflexe rozhovoru, analýzy dat a z psaní a prezentace výzkumné zprávy.*“ (Švaříček a Šed'ová, 2007, s. 160)

Sedláková (2014, s. 211) míní, že tato uměle navozená sociální interakce vzniká na popud tazatele, a coby optimální způsob získávání dat „*kombinuje standardizované i nestandardizované formy dotazování a snaží se minimalizovat jejich omezení.*“

Hendl (2005, s. 164) kontruje, že polostrukturované rozhovory jsou střední cestou při obecném vedení rozhovorů, a vyznačují se „*definovaným účelem, určitou osnovou a velkou pružností celého procesu získávání informací.*“ Taktéž vyzdvihuje dovednostní úskalí při vedení kvalitativního rozhovoru. Dodává, že na začátku dotazování je nutné prolomit případné psychické bariéry a zajistit souhlas se záznamem.

Pro zpracování polostrukturovaných rozhovorů jsem využil nástroje automatizované transkripce Whisper a UWebASR.

Rozhovory byly vedeny s účastníky výzkumného a produkčního procesu, podílejících se na vývoji technologií umělé inteligence, a to dle jejich časových možností a relevance pro tuto práci. Konkrétně šlo o projektovou manažerku Českého rozhlasu Dalii Moudrou, zástupce šéfredaktora Deníku N Lukáše Wenera, produktovou manažerku Seznam.cz Dianu Hlaváčovou a bývalého zástupce ředitele pro rozvoj Seznam Zpráv Jiřího Špačka. Pro odborný vhled do fungování technologie syntézy hlasu byla oslovena také společnost MAMA AI která se vývojem řečových technologií zabývá a spolupracuje s řadou českých médií, včetně těch, s nimiž operuje tato diplomová práce. V tomto případě byly vedeny rozhovory se zaměstnanci tohoto českého startupu Josefem Vopičkou a Jiřím Vrobelem. Všichni účastníci kvalitativního výzkumu budou představeni v následující kapitole.

## **Medailonky respondentů**

### **Dalia Moudrá (Český rozhlas)**

- Studovala bakalářský obor studium humanitní vzdělanosti na Fakultě humanitních studií UK. Šest let pracovala v ZUNO bank jako obsahová specialistka. Tři roky působila ve společnosti Seznam.cz jako projektová a produktová manažerka. Stejnou pozici zastávala i v FTV Prima a nyní v Českém rozhlasu, kde mimo jiné koordinuje vývoj syntetických hlasů.

### **Diana Hlaváčová (Seznam.cz)**

- Vystudovala podnikovou ekonomii na Vysoké škole ekonomické v Praze. Zkušenosti sbírala jako finanční analytička v UniCredit Bank a od roku 2022 pracuje ve společnosti Seznam.cz, kde nyní v pozici produktové manažerky zastřešuje vývoj velkých jazykových modelů.

### **Lukáš Werner (Deník N)**

- Vystudoval žurnalistiku a mediální studia na Univerzitě Karlově. Nejdříve pracoval jako překladatel a korektor, první novinářské zkušenosti získával v časopisu Týden. Ve své kariéře působil také na zpravodajském portálu iDnes, zpočátku coby redaktor a následně i editor zpravodajství. Odtud roku 2015 přešel do Aktuálně.cz, zpočátku na pozici editora, později se stal vedoucím domácího týmu Aktuálně.cz a Hospodářských novin. Od roku 2018 působí jako zástupce šéfredaktora v Deníku N.

### **Jiří Špaček (dříve Seznam Zprávy)**

- Na Univerzitě Karlově vystudoval sociologii a mediální studia. Pracoval na volné noze jako novinář, pak se začal zajímat o marketingovou komunikaci na internetu. Od roku 2012 pracoval na vývoji zpravodajského webu Lidovky.cz či videoplatformě Playtvak.cz. Mezi lety 2016 až 2023 působil v Českém rozhlasu, kde pracoval na vývoji a spuštění zpravodajského serveru iRozhlas.cz. Podílel se také na vývoji aplikace můjRozhlas i zavádění syntetických hlasů v Českém rozhlase. Roku 2023 přešel do redakce Seznam Zprávy, kde coby zástupce ředitele pro rozvoj měl Špaček na starosti zavádění inovativních technologií a postupů do praxe. Pod jeho vedením se zde spustily nástroje jako Tesa, Edna, Xena či Stela a stál u vývoje syntetických moderátorů jako Hacsiko. V Seznam Zprávách pracoval do února 2025.

### **Josef Vopička (MAMA AI)**

- Inženýrský titul Vopička získal z mikroelektroniky na Českém vysokém učení technickém v Praze. Zde i pokračoval v doktorském studiu, které dokončil na Katedře elektrotechniky, a již během vysokoškolských studií se specializoval na strojové rozpoznávání řeči. Jako softwarový výzkumník a vývojář pracoval od roku 1997 pro různé divize technologické společnosti IBM, především se věnoval oblasti umělé inteligence. Spolu s dalšími odborníky na strojové učení se od nadnárodní firmy oddělili, aby roku 2021 založili startup MAMA AI, kde se zaměřují na podnikovou implementaci AI technologií, jako jsou hlasoví asistenti.

### **Jiří Vrobel (MAMA AI)**

- Absolvoval obor informatiky na Fakultě elektrotechnické Českého vysokého učení technického v Praze. Největší část profesní kariéry působil v technologické společnosti IBM. Tricetileté zkušenosti z oblastí prodeje, business developmentu či datové analytiky od roku 2022 využívá ve startupu MAMA AI coby zástupce pro podnikovou implementaci a řešení umělé inteligence.

## Výzkumná část

Součástí praktické části diplomové práce je analýza uskutečněných polostrukturovaných rozhovorů s účastníky kvalitativního výzkumu. Pro maximální přehlednost jsou výpovědi uspořádány dle pracovišť respondentů. Plné znění rozhovorů je k nahlédnutí v archivu autora této diplomové práce.

## Vývoj syntézy hlasu (MAMA AI)

Pro účely této diplomové práce a specifické ukotvení tématu v českém kontextu autor oslovil odborníky na vývoj syntetizovaných hlasů ze soukromé společnosti MAMA AI, založené roku 2021 šesticí odborníků, kteří 25 let pracovali v oblasti umělé inteligence pro nadnárodní technologickou společnost IBM. MAMA AI vyvíjí vlastní hlasové asistenty, technologii vyhledávání přirozeným jazykem i telefonní aplikace v rámci dceřiné společnosti Telma AI.

Jak vysvětluje Josef Vopička, základem technologie text-to-speech je procesování textu, při němž se vstupní text upravuje tak, aby byl vhodný pro další fáze převodu na zvuk. Pomocí textové normalizace se převádějí symboly, zkratky, čísla nebo cizí slova do čitelnější podoby, takzvané verbalizace. Tokenizací se zase text rozdělí na menší celky, věty a souvětí. Ortografická povaha českého jazyka vývoj sice usnadňuje, podle Vopičky jde však stále o složitý proces. Hybridní přístup procesování textu využívá gramatiky pro generování pádů a rodů. „*Jazykový model pak vybírá nejpravděpodobnější variantu na základě statistické pravděpodobnosti, například pro Karel IV. z kombinací Karel čtvrtá, čtvrtý, čtvrtého a tak dále,*“ uvádí Vopička příklad.

Druhý krok zahrnuje převod normalizovaných slov na nejmenší části zvukové stránky lidské řeči, a to na hlásky, tedy fonémy. „*Pro tento účel se používá mezinárodní fonetická abeceda (International Phonetic Alphabet, IPA, pozn. aut.), která sdružuje a snaží se popsat všechny hlásky celého světa pomocí specifických uniodových znaků umožňující popis pozice hlasového traktu, například pozice jazyka, úst, znělosti a řady dalších fonetických detailů,*“ vysvětluje Josef Vopička s tím, že se dále řeší hlásková spodoba znělosti českého jazyka, tedy asimilace. „*Neříkáme ‚v kině‘, ale ‚f kině‘. V angličtině jsou obdobné homografy, typicky u slov ‚read‘ (přítomný a minulý čas) a ‚lead (sloveso vést nebo olovo)‘. Tato problematika se řeší pomocí rozsáhlých slovníků, z nichž lze natrénovat takzvaný sequence-to-sequence (seq2seq) model, který funguje jako hlavní mechanismus pro modelování vztahu mezi vstupní sekvencí, textem, a výstupní sekvencí obsahující*

*nejpravděpodobnější sekvenci hlásek. Model je zároveň schopen zobecňovat, to znamená domýšlet si výslovnost slov, které ani neviděl v trénovacích datech.“*

Z výsledné sekvence hlásek pak model neuronových sítí, jež jsou současným standardem v oblasti vývoje syntézy hlasu, vytvoří takzvaný mel-spektrogram. Mel spektrogram lze empiricky napočítat pro daný audio-signál. Modelu je při tréninku předkládán trénovací text a mel-spektrogram a model se tímto způsobem učí predikovat spektrogram z textu. Vokodéry, které se poté naučí z vizualizace mel-spektrogramu zrekonstruovat zvukový signál, tedy samotnou řeč, pak na základě vizuální reprezentace mel-spektrogramu vygenerují konkrétní zvuk. Model neuronových sítí pak porovnává výstupy vokodérů se skutečným zvukovým signálem. *„Jakmile vokodér vygeneruje zvuk, můžeme optimalizovat váhy, aby se výstup podobal reálnému zvukovému signálu. Dále tak můžeme upravovat například prozódii řeči – její rychlost či výšku,“* objasňuje Vopička.

Vývoj hlasové syntézy podle něj lze stále vylepšovat na několika úrovních. *„V audio je obsaženo mnohem více informací a emocí než v prostém textu. Můžeme zdokonalovat kvalitu hlasu i výslovnost neznámých slov. Nejtěžší pak bude zdokonalovat konkrétní emoce, na to je potřeba poměrně hodně trénovacích dat. S každou emoční polohou se navíc množství dat násobí,“* říká Vopička.

Technologie rozpoznávání řeči speech-to-text, která má praktické uplatnění kupříkladu u nástrojů transkripce, podle Vopičky funguje obdobným způsobem jako syntéza řeči text-to-speech. *„Zvukový signál se rovněž převede do mel-spektrogramu, který pro zjednodušení můžeme připodobnit obrázku, a když se neuronová síť trénuje a má k dispozici přepsaný text, neuronový model se snaží z tohoto empiricky vypočítaného obrázku predikovat, co je na něm napsáno,“* říká s tím, že MAMA AI pracuje s nástrojem Whisper, který vylepšuje rozpoznávat češtinu. Nástroj společnosti OpenAI Vopička hodnotí ambivalentně. *„Může být občas dokonalý, ale také dělá dost chyb. Přepisuje spíše foneticky, tudíž vyměňuje krátké hlásky za dlouhé. Snaží se totiž napodobit a vygenerovat, co člověk řekl, třeba i nespisovně. Zásadnější problém však podle mě představují zvukové nahrávky s hlukem na pozadí či místy, kde lze namísto řeči slyšet neurčitý hluk. V takovýchto chvílích Whisper halucinuje, vymýšlí si. Případně se kvůli těmto problémům může zakotkat a dokola opakuje jedno konkrétní slovo či slova.“*

Jak bude popsáno v následujících kapitolách, MAMA AI spolupracuje například se zpravodajským serverem Seznam Zprávy, pro který natrénovala pětici syntetických hlasů využívaných v denním zpravodajství, a Českým rozhlasem, přičemž implementaci řešených technologií tato diplomová práce sleduje. Základ technologie text-to-speech sice vyžaduje násobně více hodin vstupních nahrávek, pro trénování konkrétních hlasů si firma vystačí s přibližně pěti hodinami. Nahrávání probíhá v profesionálním studiu, kde lidský hlasový talent čte předpřipravené texty určitým stylem, a to na základě předchozí domluvy s klientem. Pakliže je syntetický hlas určen například pro účely audioverzí článků, obsahuje delší souvětí a složitější pojmy. Blíže tato specifika přibližují následující kapitoly s konkrétními případy využití syntézy řeči v redakcích.

Právně je využití cizího hlasu striktně ošetřeno. „*Hlas je podle českého zákona osobnostním rysem, tudíž není možné jej svévolně používat a napodobovat,*“ upozorňuje Jiří Vrobel, kolega Josefa Vopičky. „*Pokud pro někoho připravíme smlouvu, vždy se jedná buďto o trojstrannou dohodu či smlouvu, kdy náš klient dodá hlasový talent, jehož hlas bude využívat, a pak musí uzavřít smlouvu s jeho poskytovatelem,*“ vysvětluje Vrobel.

Velký prostor pro využití technologie text-to-speech podle Vopičky skýtá právě mediální odvětví. „*Lidé čím dál více konzumují mediální obsah v audio. Umožní jim to sledovat zprávy, články si pouštět do uší, navíc u toho mohou i řídit.*“ Mimo úzké zaměření kvalitativního vzorku této diplomové práce využívají služby této společnosti také veřejnoprávní Česká televize a komerční TV Prima. „*Konkrétně jde o nástroj mVoiceStudio, do kterého se nahráje film, a editor k němu napíše takzvané audiodeskripce. Jde o zvukovou stopu uvnitř filmu, kterou si mohou zapnout například slabozrací lidé, a syntetický hlas jim říká, co se zrovna děje na obrazovce,*“ popisuje Josef Vopička.

Podle Vopičky dříve mediální organizace braly syntetické hlasy s rezervou. „*Dnes ale začínají chápat, že s vývojem se technologie text-to-speech stává použitelnou pro tvorbu obsahu. Samotná syntéza však nevymyslí nic, co jí sami nepředložíme,*“ říká.

Pro syntetizování konkrétního hlasu je zapotřebí pořídit jeho nahrávku v profesionálním studiu s nejvyšší možnou kvalitou. „*Během nahrávacího dne se pořídí asi jedna hodina zvuku, zkušenější hlasové talenty zvládnou i delší nahrávání. Přestože dokážeme nasnímat hlas tak, že danou osobu poznáte z několikaminutové nahrávky, pro kvalitní hlasový model s odpovídající prozodií je potřeba kolem pěti hodin čistého audia,*“ objasňuje Josef Vopička. Po kontrole, zdali texty souhlasí

s pořízeným audiem, se dataset nahraje do neuronové sítě a začíná trénink dat. Trénování modelu trvá přibližně tři týdny.

Problematikou technologie text-to-speech je také její případné zneužití. „*Všechno se dá zneužít. Syntetické hlasy poskytujeme pro etické účely, nikoliv pro hate speech. Veškerý traffic time se nám navíc zapisuje do logů. Pokud by tedy některý ze zákazníků naše produkty zneužíval, jednoduše je vypneme,*“ vysvětluje Vopička a dodává, že tak jako si lidé vytvořili intuici při obdržení podvodných e-mailů, stejný postup budou muset uplatnit i v případě podvodů využívající syntetické hlasy.

Etický rámec využití syntézy hlasu se však nepojí význačně s jejím zneužitím. Jiří Vrobel mluví o pokusném projektu firmy, kdy pomocí umělého hlasu volali pacientům s Alzheimerovou chorobou a připomínali jim medikaci léků. (Javůrek, 2023) „*Podle pořízených dialogů se lidé k robotovi vyjadřovali, jako by hovořili k živému člověku, oslovovali ho i zdobnělinou ‚Tomášku‘. Lze tedy vidět, že navázat vztah s robotem je, bohužel, možné. Pro řadu lidí, v tomto případě šlo o důchodce, to totiž možná byl jediný kontakt s ‚lidským‘ hlasem za celý den. Když projekt skončil, dožadovali se, aby jim naše hlasy i nadále volaly.*“ Vrobel i Vopička se shodují, že nadměrná konzumace syntetických asistentů, kteří mají pomoci například introvertním či osamělým lidem, osamění naopak prohlubují. „*Samota je pro nás stresující. Nelíbí se mi představa, že nahradíme kontakt se skutečnými lidmi syntetickými hlasy a roboty,*“ dodává Jiří Vrobel.

## **Shrnutí**

V rámci sekce byl popsán princip fungování technologie syntézy řeči (text-to-speech) i rozpoznávání řeči (speech-to-text), přičemž firma MAMA AI se vývoji obou technologií věnuje. Současným standardem při jejich vývoji je využití neuronových sítí, technologii hlasové syntézy však lze vylepšovat na vícero úrovních – kvalita syntetizovaného hlasu, výslovnost neznámých slov, zvláště náročné pak je zdokonalení konkrétních emocí, k nimž je zapotřebí velký objem trénovacích dat. Rozpoznávání řeči nachází využití například u nástrojů transkripce, syntéza hlasu je využívání v rámci tvorby syntetických hlasů pro mediální organizace. Své nástroje a služby MAMA AI poskytuje českým médiím, jako jsou Seznam Zprávy, Český rozhlas, Česká televize či TV Prima.

Základ technologie syntézy hlasu vyžadují mnoho hodin audionahrávek, konkrétní modely je však firma schopna syntetizovat z pěti hodin audiozáznamů pořízených ve studiové kvalitě, kde hlasový

talent předčítá připravené texty určitým stylem na základě předchozí domluvy s klientem. Trénování jednotlivých modelů trvá firma přibližně tři týdny. Právní využití cizího hlasu je ze strany MAMA AI striktně ošetřeno, a podléhá etickým účelům. V případě zneužití hlasových modelů MAMA AI ze strany klienta je společnost schopna sledovat jejich využití a v nejzazším případě je i vypnout. Respondenti však nabádají k opatrnosti, nadměrné využívání syntetických asistentů pro osamělé či introvertní osoby podle nich naopak osamění prohlubuje.

## Deník N

*„Technologie text-to-speech je pro nás samozřejmostí,“ otevírá rozhovor zástupce šéfredaktora Deníku N Lukáš Werner. „Deník N je typický delšími texty a řada čtenářů, i bývalých, opakuje, že nemá na čtení čas. Kvalitně zpracovaná syntéza hlasu je proto způsob, jak lze obsah pohodlně konzumovat.“*

S myšlenkou načítaných textů přišli zástupci Deníku N již před několika lety. Technologie syntézy řeči tehdy ještě nebyly na takové úrovni jako dnes, a proto se zpravodajský server vydal cestou lidských předčítačů. Zpočátku server oslovil moderátory z rádia. *„Spolupráce však byla ukončena po řadě problémů. Moderátoři byli zvyklí, že jim v rádiu texty přepsali foneticky, a to včetně všech jmen. Nebyli schopni samostatné práce a obraceli se na editory s požadavky na vyhledávání správné výslovnosti řady jmen, i notoricky známých. Jedna z moderátorek například přečetla ‚ruřina‘. Pro editory to bylo značně zatěžující. Neměli jsme kapacity po nich každou nahrávku kontrolovat, chyby jsme vychytávali až na webu. A občas byly poměrně zásadní.“*

Deník N poté navázal spolupráci se skupinou studentů herectví. *„Řada z nich byla velmi talentovaná, proto posléze načítali i extrémně dlouhé články – například dějiny Izraele. Čtenáři si takové formáty oblíbili, bohužel jsme ale z časových a finančních důvodů nemohli načítat více než čtyři až pět článků denně.“* Právě na nízký počet načtených textů si podle Wenera čtenáři stěžovali nejvíce. Výtky přicházely taktéž na preferenci jednotlivých hlasů. *„Někomu vadil mužský, někomu zase ženský hlas,“* vysvětluje Werner.

V Deníku N, jehož financování funguje na základě modelu platební brány, vyzkoušeli i hybridní model, kdy podle Lukáše Wenera zavedli *„velmi nedokonalého robota“*. Ten načtel zbytek článků, které nestihl přečíst lidský hlas. *„Nakonec to však dopadlo, jak muselo – zatímco se robot neustále zlepšoval, lidé zůstávali stejní, a byli nahrazeni nesrovnatelně levnějším a rychlejším strojem.“*

Ne však nutně kvalitnějším. *„Spousta lidí je ochotna přistoupit na kompromis, kdy robot sice není dokonalý kvůli chybám v intonaci, výslovnosti či robotickém hlase, nicméně je k dispozici okamžitě ve všech textech,“* míní Werner s tím, že od syntetických hlasů již není cesty zpět.

Deník N také využívá technologii speech-to-text v rámci softwaru společnosti Beey.io k transkripci rozhovorů. *„Je to pomocník. Doufáte, že přepis dopadne dobře a software text správně přepíše.“*

*Roli hraje také prostředí, v němž nahrávka vzniká. Rušná kavárna zapříčiní oproti tichému studiu úplně jiné výsledky. Jindy se však v přepisu objevují nesmysly. Čas tato technologie nicméně šetří, jsou ale mezi námi redaktori, kteří si to nemyslí. Takoví lidé jsou schopni psát deseti prsty a rovnou při poslechu nahrávky vyškrtávat slovní vatu.“* Werner usuzuje, že je momentálně běžné, aby transkriptor přepsal slovo od slova a na každém redaktorovi pak je tisíce přepsaných znaků upravit. Byť jde podle něj o velmi užitečné nástroje, práci novináře momentálně nedokáží nahradit. *„Stroj mi sice zajistí přepis, ale stejně nemám ani z poloviny hotovo.“*

Zpravodajský server je zároveň závislý na vývojářských kapacitách slovenského Denníku N. Když se v Česku rozhodli o zapojení robotického hlasu do svého provozu, využili toho, že systém už měli vývojáři připravený pro slovenský Denník N. Zbývalo jen najít vhodné hlasy trénované na českém jazyce a ty doladit. Werner zároveň připouští, že Deník N od svého počátku v roce 2018 technologickou stránku rozvoje online média podcenil, neboť pro ně zprvu nebyla podstatná. *„Vždy jsme se soustředili především na to, aby u nás lidé našli dostatek unikátního obsahu. Vedle jiných médií, například Seznam Zpráv, jsme malincí hráči. Produkce článků od roku založení narostla, ale technologické základy našeho webu zůstaly stejné,“* prohazuje. V Deníku N se proto rozhodli do produkčního procesu zapojit nástroj Azure od společnosti Microsoft. Werner vyzdvihuje jeho výhody. *„Především cena za syntézu byla velmi příjemná. Syntetizovat jeden článek totiž stálo pár haléřů. Měl však limity. Články s počtem znaků vyšší než několik desítek tisíc nezvládl syntetizovat. To však možná byla chyba našeho Wordpressu (redakčního systému, pozn. aut.). Tento problém jsme ale řešili velice sporadicky,“* dodává.

Deník N pro účely syntézy hlasu v rámci Azure využíval dva hlasy – Antonína a Vlastu. Werner doplňuje, že stejné hlasy využívá například také bankovní či pojišťovnický sektor u svých zákaznických linek. S volnou přístupností těchto hlasů se však pojí také riziko šíření dezinformací. Jedna z redaktorek Deníku N se dle Wenera stává opakovaným cílem podvodných videí, na kterých je sice skutečná redaktorka, má však hlas syntetického hlasu Vlasty. Stejně jako každá jiná technologie je i tato zneužitelná.

Řečové technologie však Deník N neimplementoval bez předchozího nastavení jejich parametrů. *„Jsou to však kosmetické úpravy. Jeden hlas byl defaultně pomalejší, a proto jsme jej mírně zrychlili, aby mezi tempem mužského a ženského hlasu nebyly tak velké rozdíly. Zároveň jsme se snažili řešit i jeho poněkud neohebnou intonaci zejména na konci vět. To znamená, aby se hlas jinak zachoval, když má na konci věty tečku, nebo naopak otazník. To se nám už ale příliš*

*nepovedlo. Jestliže tento proces chcete automatizovat, musíte vytvořit pravidlo typu ‚Když je otázka postavená takto, jde hlas nahoru, když jinak, jde hlas dolů.‘ Tak daleko jsme se ale nedostali, vyžadovalo by to větší zásahy vývojářů a kdoví, jestli by to tomu vůbec pomohlo,“* kroutí hlavou zástupce šéfredaktora.

Limitované schopnosti projevily syntetické hlasy zpočátku dokonce i při vyslovování jmen autorů článků. Příjmení Jana Wirnitzera robotický hlas přečetl jako „*Virnicer*“. Navzdory kvalitní úrovni českého jazyka se software Azure potýkal s problémy při vyslovování přejímaných slov.

*„Přejatá slova typu ‚fastfood‘ byla nejhorší,“* objasňuje Werner a pokračuje. *„První pád slova Azure zvládal solidně, nicméně například frázi ‚do fastfoodu‘ přečetl jako ‚do fastfódu‘. Měli jsme proto primitivní postup řešení. Do sdílené tabulky jsme foneticky prepisovali výrazy, jak je má robotický hlas vyslovit. Znamenalo to však povinnost fonetický zápis přidat ke každému pádu a tvaru daného slova, a to bylo neuvěřitelně ubíjející.“* Kromě jmen a přejímaných slov má Azure obtíže s řadovými číslovkami. Není kupříkladu schopen správně přečíst „*v 19. století*.“

Funkčnost softwaru však začaly provázet komplikace neslučitelné s efektivním využíváním. *„Na podzim 2024 Microsoft asi nahrál novou verzi modelu pro náš region, přičemž zapomněl optimalizovat slabikotvorné ‚r‘ a ‚l‘,“* usuzuje Werner. *„Robotický hlas pak zněl jako osoba s poruchou řeči vynechávající ‚r‘. Do toho se objevily i jiné, hůře vysvětlitelné jevy: slovo ‚Vltava‘ robot přečetl jako ‚Vstava‘. Problém jsme sice nakonec vyřešili změnou lokality mimo střední Evropu, od té chvíle jsme však začali hledat jiná řešení.“* Dodává, že nynější systém je na svém maximu a již ho nelze výrazně zlepšovat.

Alternativu proto Deník N našel ve službě polské společnosti ElevenLabs, která se na syntézu řeči specializuje. A podle Wenera má tato firma značný náskok před Microsoftem. Rozhodující v tomto ohledu byla multilingválnost služeb ElevenLabs, jejichž hlasové modely disponují češtinou, slovenštinou i maďarštinou, dále možnost vytvořit si vlastní syntetické hlasy a také finanční náklady za službu. Vzhledem k propojenosti redakčních systémů českého Deníku N, slovenského Denníku N a maďarského Napunk se online média nakonec rozhodla pro služby ElevenLabs. Firma má ve své nabídce desítky natrénovaných hlasů určených pro různé oblasti, včetně předčítání zpravodajského obsahu. V rámci pořízeného tarifu budou mít vydavatelství společný účet. ElevenLabs má oproti nástroji Azure specifikum v opravách generovaného textu. *„Často se texty v čase vyvíjejí a dopisují se k nim dodatečné informace, proto je záhodno mít*

*aktualizovanou audioverzi článku. U nástroje Azure se musí přegenerovat celý text, kdežto ElevenLabs umí přegenerovat i jeden konkrétní odstavec. Díky tomu ušetříme spoustu syntetizovaných znaků,*“ vyzdvihuje Lukáš Werner.

Součástí řešení bude také již zmíněná možnost tvorby vlastních syntetických hlasů. Deník N chce, mimo jiné pro potřeby rozhovorů, syntetizovat hlasy čtyř hlasových talentů – dva ženské a dva mužské, a modely bude trénovat na vlastních autorských textech, které jsou typické svým delším rozsahem.

V době psaní této diplomové práce Deník N již syntetizoval hlas šéfredaktora Deníku N Pavla Tomáška. *„ElevenLabs zmiňuje na různých místech dokumentace odlišné informace, co se týká délky audionahrávek pro trénování. Někde doporučuje dvouhodinové audiozáznamy, jinde hovoří o třech. Protože jsme brali hlas Pavla Tomáška jako takovou první zkoušku schopností, a zároveň jsme ho nechtěli zbytečně okrádat o jeho čas tím, že by seděl zavřený ve studiu déle, než by bylo potřeba, přiklonili jsme se ke spodní variantě, a nahráli jsme tudíž něco málo přes dvě hodiny. I systém ElevenLabs nás upozorňoval na malé množství materiálu. Výsledek je ale více než přesvědčivý. Je možné, že Tomáškův hlas v budoucnu upgradujeme. S dalšími třemi hlasy však budeme cílit na tříhodinové nahrávky,*“ říká Werner. Vedle Pavla Tomáška Deník N syntetizuje také hlasy redaktorek Lenky Vrtiškové Nejezchlebové, Jany Ciglerové a zástupce šéfredaktora Michaela Švece, a pokud se technologie osvědčí, bude chtít mít k dispozici i širší paletu hlasů z vlastní redakce.

K syntetizovaným audionahrávkám plánuje Deník N přidávat i upozornění, že hlasy nejsou reálné. O konkrétní podobě však ještě nepadlo konečné rozhodnutí, stejně jako o smlouvách, které Deník N plánuje se svými zaměstnanci poskytující hlasy uzavřít. *„Jsme domluveni, že poskytnutí hlasu je bez nároku na honorář. Vážím si, že do toho jdou. K nasazení syntézy do provozu pak sepíšeme jednoduchou smlouvu, čistě formální. V případě externistů bychom však sestavili smlouvu obsahující různé scénáře, které by mohly v průběhu spolupráce nastat,*“ dodává Werner.

Využití syntetických hlasů v české mediální krajině podle Lukáše Wernera do značné míry determinuje postavení technologické společnosti Seznam.cz. *„Čtenáři nám občas vyčítají, že naše syntetické hlasy jsou výrazně horší než umělé hlasy na Seznam Zprávách, které své články zpřístupňují zadarmo všem čtenářům. To je sice pravda, jenomže Seznam je technologická firma, v Deníku N si nemůžeme dovolit vlastní tým vývojářů a klást takový důraz na vlastní vývoj*

*technologických novinek,*“ kontruje Werner. Stále však i přes nedokonalosti považuje dosavadní využití technologie syntézy hlasu za významné zkvalitnění obsahu, jenž umělé hlasy činí přístupnějším cílovému publiku, a kvalitativní laťka – byť za cenu nemalých nákladů – by se měla zvyšovat během letošního jara.

## **Shrnutí**

Deník N vnímá implementaci syntézy hlasu jako prostředek k rozšíření dostupnosti svého obsahu, zejména s ohledem na časově vytížené čtenáře. Po počátečních experimentech s lidskými předčítači a hybridním modelem přešel redakční tým postupně na plně automatizovanou syntézu, která je levnější a efektivnější, přestože kvalita stále zaostávala za lidským přednesem. Využívání syntetických hlasů je považováno za nevratný krok, i přes přetrvávající nedostatky v intonaci či výslovnosti.

Pro zpracování syntetických hlasů Deník N původně využíval službu Azure od Microsoftu, která se vyznačovala nízkými náklady, avšak potýkala se s technickými omezeními a problémy s výslovností. V reakci na to redakce přešla k společnosti ElevenLabs, jež nabízí pokročilejší modely, multilingvální podporu a možnost syntetizovat vlastní hlasové talenty. Do svého provozu plánuje Deník N zařadit dva ženské a dva mužské hlasy – konkrétně hlasy Lenky Vrtiškové Nejezchlebové, Jany Ciglerové, Pavla Tomáška a Michaela Švece. Syntetické hlasy tak zůstanou klíčovým nástrojem pro zlepšení přístupnosti obsahu, přičemž redakce sleduje vývoj těchto technologií a zvažuje jejich optimalizaci.

## Český rozhlas

Veřejnoprávní Český rozhlas nasadil technologii syntézy hlasu do ostrého provozu svého webu iRozhlas.cz a dalších platformem v květnu roku 2024. Řečové syntéze předcházelo interní výběrové řízení, do kterého se mohli zapojit všichni zaměstnanci a spolupracovníci Českého rozhlasu. Z více než třiceti přihlášených osob porota vybrala hlas moderátora Pavla Prouzy. Po interním testování Rozhlas nasadil do provozu i druhý, ženský hlas. Stále však probíhá jeho optimalizace a identitu ženského hlasového talentu Český rozhlas veřejně nekomunikuje.

Produktová manažerka Českého rozhlasu Dalia Moudrá vysvětluje, že syntézu hlasu Český rozhlas využívá především pro účely zvukových verzí článků. *„Coby instituce pracující s audiem se držíme základního pravidla. Pokud máme k danému článku k dispozici skutečné audio z živého vysílání, vždy jej doplňujeme touto nahrávkou.“* V opačném případě se rozhoduje, zdali ke článku vytvořit syntetickou audioverzi. Zpočátku zaměstnanci Rozhlasu takto vybírali jeden až dva články denně. V průběhu uplynulého roku se číslo mírně zvýšilo na průměrně pět syntetizovaných článků za den. Z celkově sedmdesáti až osmdesáti průměrně vydaných textů na zpravodajském serveru iRozhlas.cz jde tedy o malé procento textů. *„Zhruba polovina článků má nicméně dostupné audio z živého vysílání, tam tedy potřeba umělých hlasů odpadá. Výběr zbylých textů spadá na konkrétního člena redakce, který má syntézu hlasu v kompetenci.“*

Redakce prozatím nemá jasně stanovená konkrétní pravidla pro výběr článků určených k řečové syntéze. Při personálních změnách na pozici spravující hlasovou syntézu tak dochází i k výběru nepůvodních textů zpravodajského serveru iRozhlas.cz. Dle názoru Moudré však redakce směřuje k definování jasnějších pravidel.

Mediální instituce si již nicméně stanovila podrobná pravidla pro práci s umělou inteligencí, a to včetně technologie hlasové syntézy. (Český rozhlas, 2024)

V souvislosti s těmito směrnicemi Dalia Moudrá hovoří o robustním informačním základu sestaveném pro způsoby, jakými Český rozhlas přistupuje k nástrojům uplatňujícím algoritmy, strojové učení a umělou inteligenci. *„V rámci pracovní skupiny, která se na tvorbě těchto směrnic pro Český rozhlas podílela, jsme se shodli, že DNA Českého rozhlasu spočívá v tvorbě kvalitního audio obsahu. Vytváříme ho ve spolupráci s kvalitními moderátory, interprety a herci. Nástroje umělé inteligence používáme spíše pro věci, které bychom jinak nedělali, a umožní nám vytvářet*

*nadstavbu.*“ Dodává, že mediální organizace rozhodně nesměruje k nahrazování lidských moderátorů těmi umělými.

*„V tuto chvíli rozhodně nejednáme o nasazení syntézy do vysílání. Vždy se snažíme tyto nástroje využívat na aktivity, které bychom normálně nedělali. Vzhledem ke kapacitám bychom určitě nenamlouvali články živými lidmi. Zároveň však chceme našim uživatelům poskytnout možnost si články přehrát i v audio.“*

S každým hlasovým talentem ČRo uzavírá smlouvu, ve které užití hlasu specifikuje. Osoba poskytující svůj řečový projev nemá možnost ovlivnit, pro syntézu jakých textů bude její hlas použit, samotné užití se však řídí kodexem ČRo. Součástí smlouvy je dle Moudré také finanční plnění ve vztahu k hlasovému talentu. Tyto dohody už nejsou veřejné, neboť jde o smlouvy s fyzickými osobami. *„Znamená to, že hlasový talent nám propůjčuje svoji licenci a my jsme oprávněni hlas používat. Smlouva také stanovuje dobu, po kterou můžeme hlas využívat,“* uvádí Moudrá.

Na tvorbě mužského hlasu Pavla Prouzy a ženského hlasu spolupracoval Český rozhlas v rámci veřejné zakázky s českou firmou SpeechTech. Ta si pro vývoj umělého hlasu připravila zpravodajské texty ze serveru iRozhlas.cz, které odpovídaly zhruba dvaceti hodinám řečového projevu. Moderátor Prouza podle svých slov dokázal texty číst přibližně hodinu v kuse a nahrávání trvalo asi tři měsíce. Jak v dokumentu Českého rozhlasu o umělé inteligenci *Ve službách Českého rozhlasu* (Géla a Moravec, 2024) zmiňuje Jindřich Matoušek z Katedry kybernetiky Západočeské univerzity v Plzni, s větším množstvím dat je možné dojít k lepšímu výsledku. *„Pokud bychom chtěli syntetický hlas Pavla Prouzy využít například pro čtení beletrie, v současné době by to už nevyznělo tak pěkně. Čtení beletrie vyžaduje jiný styl řeči, než je čtení zpráv.“*

Veřejnoprávní instituce je mimo jiné také součástí grantového projektu vyhlašováním Technologickou agenturou ČR, v rámci něhož spolupracuje s firmou MAMA AI. Firma MAMA AI proto vytvořila další ženský hlas, určený nikoliv pro namlouvání článků, ale výhradně k výzkumným účelům daného grantového projektu, jehož cílem je vytvořit audio asistenta, který uživatelům umožní větší míru interakce a personalizace obsahu. Ta může mimo jiné v budoucnosti Českého rozhlasu a jeho platformem hrát důležitou roli.

*„Představa je taková, že například v mobilních aplikacích, ať už aplikaci iRozhlas nebo mujRozhlas, bude implementován hlasový asistent doporučující uživateli nové epizody jeho*

*oblíbeného pořadu či nabízející shrnutí zpravodajských událostí daného dne,*“ vysvětluje Dalia Moudrá s tím, že prototyp audio asistenta by měl vzniknout v průběhu let 2025 a 2026.

Stejně jako Deník N také Český rozhlas vývoj syntézy hlasu outsourcuje na externí firmy, a spolupracuje se zmíněnými firmami SpeechTech a MAMA AI. *„V Rozhlasu nemáme ambici systémově zavádět AI technologie, které bychom si vyvíjeli sami, z důvodu, že know-how mají vždy jiní. Zato naše odbornost spočívá v audio. Co se nám však velmi osvědčilo, je interní experimentování s různými nástroji AI ještě předtím, než na ně vypíšeme veřejnou zakázku za účelem systémového užití. Chceme si totiž s předstihem vyzkoušet, zdali nám využití daného nástroje či technologie dává smysl, nebo ne. Většinou ale využíváme veřejně dostupné služby typu OpenAI a napojujeme se na jejich API a developerské přístupy.*“ objasňuje Dalia Moudrá s tím, že pokud se interní týmy ČRo nástroje rozhodnou systémově implementovat, následuje oficiální vypsaní veřejné zakázky pro firmy dodávající takovéto služby na trhu.

Český rozhlas audioverze článků označuje vizuálním piktogramem. Moudrá doplňuje, že ačkoli samotný zvuk upozornění neobsahuje, tato skutečnost se vzhledem ke zpětné vazbě, například nevidomých uživatelů, kteří o pravosti hlasů pochybují, v blízké budoucnosti nejspíše změní a upozornění bude zahrnuto do zvukové stopy.

Mediální instituce zároveň využívá automatizovanou transkripci živého vysílání. Moudrá podotýká, přepsaný text nyní slouží Rozhlasu jak pro interní účely a kategorizaci obsahu, tak i pro uživatele, kteří mají možnost přepis nalézt na webu Českého rozhlasu.

Svůj nástroj Českému rozhlasu poskytovala rovněž v rámci veřejné zakázky firma SpeechTech. Začátkem tohoto roku ČRo uzavřel smlouvu se společností Newton technologies vyvíjející nástroj Beey.io. *„Cílem je používat jejich systém pomocí API tak, aby se našim uživatelům vůbec nezměnil interface webové stránky, na který byli doteď zvyklí, ale stále dostali přepis od Beey.io.*“

Konkurenceschopnost Českého rozhlasu spočívá dle názoru Dali Moudré v jeho veřejnoprávní úrovni. *„Děláme veřejnou službu, která by měla posluchačům přinášet určitou hodnotu. Audio netvoříme pro vygenerování času, který můžeme vyplnit reklamou. Tím neříkám, že syntetické vysílání nemůže přinášet hodnotu, zajisté může. Avšak na tento pohled navazuje řada argumentů, proč nejít cestou syntetického vysílání a proč si naopak držet naši dosavadní kvalitu, na kterou jsou posluchači a koncesionáři zvyklí. Argument, že všichni ostatní syntézu hlasu ve vysílání používají, pro nás není podstatný. Podotýkám ale, že nemohu mluvit za Český rozhlas v horizontu*

*daleké budoucnosti,*“ dodává Moudrá, podle níž by Rozhlas případné změny otevřeně komunikoval a zdůvodňoval své argumenty.

## **Shrnutí**

Český rozhlas zavedl technologii syntézy hlasu v květnu 2024, přičemž primární využití nalezla při tvorbě audiočlánků. Redakce upřednostňuje autentické nahrávky z živého vysílání a syntetizované verze používá pouze tam, kde skutečné audio není k dispozici. Proces výběru článků k syntéze je zatím spíše ad hoc, bez jasně stanovených pravidel, a odpovědnost za rozhodování leží na konkrétních členech redakce. Mediální instituce se staví k syntetickým hlasům jako k doplňkovému nástroji, jenž má rozšířit možnosti audioobsahu, nikoliv nahrazovat lidské moderátory.

Pro zpracování syntetických hlasů Český rozhlas spolupracuje s firmami SpeechTech a MAMA AI. SpeechTech se podílel na vývoji hlasu moderátora Pavla Prouzy a probíhá testování druhého, ženského hlasu, jehož identitu Český rozhlas veřejně nekomunikuje. MAMA AI vyvinula ženský hlas pro výzkumné účely, zejména pro budoucí vývoj audio asistenta, který by měl zlepšit personalizaci obsahu. Implementace syntézy hlasu se opírá o pevné směrnice pro využívání umělé inteligence, reflektující závazek k udržení kvality audio obsahu. Do budoucna Český rozhlas plánuje rozšířit možnosti interakce s obsahem prostřednictvím hlasových asistentů, přičemž zachování veřejnoprávních hodnot zůstává prioritou.

## Seznam.cz

První syntetický hlas představil internetový portál Seznam.cz v červnu roku 2023. Od března téhož roku na jeho vývoji pracoval tým Seznam Zpráv a šlo o první syntetický hlas české moderátorky s přezdívkou Hacsiko, přičemž skutečný hlas vývojářům poskytla Barbara Hacsí, moderátorka Rádia Expres FM. Podle někdejšího zástupce ředitele pro AI v redakci Seznam Zprávy Jiřího Špačka však nebylo cílem vyvinout konkrétně syntetický hlas, ale prostředek ke kvalitnější distribuci obsahu. (Kasík, 2023)

Již od první poloviny roku 2023 však na Seznam Zprávách načítala audioverze článků robotka Jana, jejíž hlas mediální organizaci poskytl startup MAMA AI. To však vyžadovalo práci editorů, kteří nově museli uzpůsobit texty schopnostem syntetického hlasu vyslovovat určitá slova. (Kučerová, 2024) Redakce se tudíž rozhodla pro vlastní potřeby vyvinout své syntetické hlasy.

Už v době uvedení Hacsiko pracoval tým Seznam Zpráv na dalších čtyřech, respektive šesti syntetických hlasech, neboť z každého hlasu lze vytvořit více modelů. Své hlasy propůjčili novináři a novinářky Silvie Friedmannová, Magdaléna Korcová, Pavel Vondra a Jaroslav Gavenda. Všech pět hlasů bylo vyvinuto ve spolupráci s MAMA AI.

Hlasy čtyř novinářů se používají pro načítání článků na webech Seznam Zprávy, Proženy.cz a Garáž.cz, a to od prosince 2023. Dva modely čtou od března 2024 všechny zpravodajské relace na rádiu Expres FM a jeho sesterské stanici Classic Praha. Konkrétně jde o hlasy Magdalény Korcové a Jaroslava Gavendy.

Syntetické hlasy pro čtení zpráv v rádiích jsou pro takovýto účel upraveny. Zatímco rádiové zpravodajství zpravidla využívá krátké věty, psané články obsahují komplikovaná souvětí doplněná o cizí a komplexní slova. „*Tempo a dikce syntetických hlasů jsou odlišné,*“ podotýká Špaček. „*V rychlém rádiovém zpravodajství se musíte složitým větám vyhnout. Není obtížné dnes využít kdejaký hlas pro načtení zprávy nebo článku. A sice se z výsledku můžete radovat, jenže pokud hlas neodladíte pro cílového posluchače, který očekává čtený článek a nikoliv čtenou knihu, kdo bude takové výstupy poslouchat?*“ ptá se Špaček a pokračuje: „*Když použijete na zprávu ze sportu či politiky nesprávný hlas, který bude číst nesprávným tempem a špatnou intonací, může se stát, že sportovní zprávu přečte jako pohádku. Zprávu z politiky by automaticky zabarvil, a na to je třeba dát si pozor,*“ říká s tím, že hlasy pro účely Seznam Zpráv jsou od této komplikace

oproštěné. „*A i když vám mohou připadat nudné, splňují přesně to, co splňovat mají – nezabarvují sdělení. Slyšíte, co máte slyšet.*“

Nahrávání hlasů Silvie Friedmannové, Magdaleny Korcové, Pavla Vondry a Jaroslava Gavendy trvalo dva měsíce a dohromady bylo pořízeno na šedesát hodin nahrávek. Technologický proces hlasové syntézy popisuje příslušná kapitola *Vývoj syntézy hlasu*. Jak nicméně podotýká Špaček, fundament, na kterém systém syntézy řeči staví své fungování, vyžaduje mnohem více hodin zvukových nahrávek.

„*Vývoj vlastního modelu zabere opravdu spoustu času. Profesionál je schopen za hodinu namluvit 25 až 30 minut textu. Naopak neprofesionální za stejnou dobu namluví třeba pět až 10 minut. Je tedy výhodnější si najít profesionály. Naše nahrávání bylo hodně intenzivní, protože každý hlas mohl číst maximálně pouze hodinu. Navíc načítání je třeba zorganizovat pro každý hlasový talent pokud možno každý den a ve stejný čas, neboť se vám hlas mění během dne. Celou dobu musel udržet pozornost zvukař, technik i já,*“ vzpomíná Špaček.

Během nahrávání je nutné kontrolovat čtení včetně správné intonace a vědomých mezer ve větách. „*K natrénování hlasového modelu je třeba získat pět až deset hodin nahrávek čistého času,*“ míní Špaček s tím, že pro čtení byly vybrány konkrétní články, které bylo třeba předčítat určitým stylem. Texty však obsahují komplexní věty, jejichž součástí jsou složitá cizí slova nebo číslovky. „*Texty nebyly napsány pro čtení nahlas v rádiu, ale aby je lidé četli očima. Jsou zkrátka jinak strukturované. Věty určené pro rádio jsou kratší, aby se daly udýchat. I profesionálovi dá zabrat přečíst ekonomický text o inflaci se spoustou číslovek a desetinných čárek spolu s cizími výrazy. Je běžné, že hlasový talent začne jednotlivá slova přeskakovat, prohazovat či je zaměňovat. V případě chyb se daný úsek buďto zopakuje nebo se přepíše,*“ říká Jiří Špaček.

Po nasbírání dostatečného množství zvukového materiálu a jeho kontrole proces trénování hlasu převezme strojové učení, které v nahrávkách na základě pravděpodobnosti správně odhadne následující zvuk. Podle Špačka může trénink trvat i déle než 24 hodin. „*Záleží na objemu dat a technice. Tímto procesem však vytvoříte několik modelů, každý s jiným nastavením, a zjišťujete, jak je pro vaše potřeby ideálně nakombinovat,*“ míní.

Zjara roku 2024 se vedení Seznamu rozhodlo vyvinout vlastní hlas pomocí technologií nezávislých na outsourcingu. „*Seznam nyní pracuje s daty, na kterých jsou natrénované naše hlasy v rámci*

*Seznam Zpráv*, “ objasňuje Jiří Špaček s tím, že konkrétně jde o více než 11 tisíc žurnalistických článků vydaných v rámci společnosti Seznam.

Diana Hlaváčová, produktová manažerka Seznamu zastřešující vývoj velkých jazykových modelů, říká, že Seznam v budoucnu implementuje řečové technologie také do jiných sfér, a postupně rozšíří portfolio svých hlasů a produktů. Společnost podle ní vyvíjí vlastní technologie, neboť využití vidí i mimo obsahovou platformu. Konkrétně u ní pak plánuje poskytovat možnost poslechu článků v kompletní šířce svého portfolia. „*Uživateli tak přineseme ještě lepší zážitek a můžeme mu ulehčit použití našich služeb,*“ věří Hlaváčová.

Hlas syntetické moderátorky Hacsiko na rádiu Expres FM poprvé zazněl 21. června 2023, a po půl roce testování v rámci nočního vysílání bylo oznámeno i její nasazení do pořadu Trendy, zaměřeného na novinky v technologiích. (Seznam Zprávy, 2023)

V případě moderátorky Hacsiko jsou posluchači vždy na umělý hlas upozorněni. Zpravodajské relace na rádiích Expres FM a Classic Praha, kde zaznívají hlasy Magdalény Korcové a Jaroslava Gavendy, už však upozornění neobsahují. Špaček argumentuje, že v případě pouhého čtení textu je takové upozornění zbytečné. „*Při uveřejnění proběhla tisková konference. Na syntetické zpravodajství neupozorňujeme, protože je to zbytečné, hlasy pouze čtou texty připravené lidmi, žádné zprávy jako takové samy nevyrobí.*“ V případě čtení audioverzí textů byl vydán článek upozorňující na nový formát Čtete vám Seznam Zprávy. (Kučerová, 2024)

Diana Hlaváčová, připomíná diskuze o natolik realistických umělých hlasech, že pro čtenáře nebude možné je odlišit od skutečných. „*Budeme nad rámec označování, že jde o AI produkt, kvalitu záměrně kazit různými artefakty? Kdo ví. Jde nám o užitečnost pro uživatele, ani ne tak o formu. Myslím si, že v žurnalistice, ale i jiných oborech, toho lze řečovými technologiemi ještě spoustu usnadnit a zpříjemnit,*“ tvrdí Hlaváčová.

Spolu s rozvojem řečových technologií proběhl i vývoj tří systémů a nástrojů určených pro práci s jednotlivými implementacemi automatizovaných technologií:

- **Xena** je redakční systém využíváný pro generování zpravodajských relací na rádiích Expres FM a Classic Praha. Editor do Xeny vloží konkrétní zprávy v textové podobě, zařadí tyto vstupy do relace a otestuje jejich načtení. Xena poté již k relaci automatizovaně přiřadí znělku, zařadí do vysílání a obsah příslušnými hlasy v daný čas odvysílá.

- **Edna** je redakční systém pro správu vysílání a vstupů syntetické moderátorky Hacsiko. Editor vysílání může i za použití chatbotů pracovat se vstupy pro vysílání, kontrolovat jejich audioverze a zahrnout je do playlistu.
- **Stela** je redakční systém určený pro generování audiočlánků. Po vložení URL adresy článku Stela obsah oprostí od nepotřebných elementů a automaticky jej upraví pro generování syntetickým hlasem. Editor audioobsahu poté následně audiočlánek zkontroluje, vygeneruje a zařadí do distribuce. (Seznam Zprávy, 2024)

Pro další využití má Seznam připravené i autonomní agenty-moderátory. „*Jsou připravené prototypové verze třech moderátorů, dvojice a jednotlivce. Zatímco jednotlivec by vysílal sám, dvojice je určená pro konverzaci o tématu článku, který jim poskytnete, a v textu si najdou zajímavý prvek, o němž lze mluvit. Článek tyto agenti rozpracují mezi sebou do dialogu a výsledný text posléze hlasy přečtou,*“ popisuje princip syntetických moderátorů Špaček. Vývoj takovýchto agentů není podle Špačka jednoduchý. Je zapotřebí, aby umělý moderátor byl konzistentní a navazoval na předchozí dialogy. Do konverzace se navíc musí propisovat charakterové rysy moderátorů. Poté, co AI dramaturg vysílání vloží do vysílacích slotů vstupy na základě vybraných článků, autonomní robot zpracovává poskytnutá data do vysílacího vstupu, který je dále doplněn o syntetický hlas a znělky. Poté se celý vstup vloží do plánu vysílání a obsah se odvysílá.

Podle Špačka se v rámci mediálního domu syntetický hlas nepřidal, aby zvýšil konzumaci vydávaného obsahu, nýbrž aby se co nejširšímu publiku dostalo jednoho obsahu v různých formátech. „*Čtenáři i posluchači se často proměňují, a na to je potřeba rychle reagovat. TikTok či YouTube shorts kralují videím. Jestliže do takového formátu máte přidat zpravodajství, musíte se dle toho přizpůsobit. Nemá smysl jít do tvorby takového formátu jen napůl, jinak si tím zkazíte jméno,*“ říká Špaček. V mediálním domě Jiří Špaček navrhl také nástroj Tesa, který na základě článku vytvoří dočasný příspěvek na sociální síť, takzvané story.

Diana Hlaváčová míní, že řečové technologie zatím překonávají své limity. „*Přinášejí audioverze našeho oblíbeného obsahu tam, kde bychom nenašli čas si jej přečíst. V budoucnosti to třeba bude umožněné i s výběrem hlasu či rychlosti přehrávání.*“ Podle Hlaváčové chce Seznam využít AI k demokratizaci kvalitního obsahu. „*Firma Seznam.cz má mnoho zaměstnanců i médií. Bez použití syntetických hlasů by nebylo možné vytvořit dostatek nahrávek včas. Kým pak jsme, abychom vybírali, co má, nebo nemá právo být i v audioverzi?*“ ptá se.

Seznam.cz v rámci vývoje svých prvních AI produktů a jejich uživatelského testování spolupracoval s MAMA AI, ale testoval a využíval na své mediální platformě také audiotechnologie jiných firem. „V Seznamu se však už delší dobu věnujeme umělé inteligenci. Děláme tak z důvodu bezpečnosti, garance stability řešení v naší režii a též možnosti škálovat podle našich představ. Během vývoje velkých jazykových modelů jsme se na konci roku 2024 dostali k výzkumu řečových technologií. V roce 2025 plánujeme technologie třetích stran nahrazovat vlastními, a tím i otevřít možnost jednotného škálování napříč našimi médii a budovat širší škálu audioproduktů v Seznam.cz,“ doplňuje Hlaváčová budoucí vývoj.

Společnosti Seznam.cz řečové technologie také výrazně navyšují reklamní prostor. Mediální organizace díky nim může rozšířit způsoby monetizace i cílit audio reklamu. (Seznam Zprávy, 2024)

Čtveřici novinářských hlasů lze slyšet na webech Seznam Zprávy, Proženy.cz a Garáž.cz. Základ článků namlouvají obvykle syntetické hlasy Silvie Friedmanové a Pavla Vondry, při užití citací lze slyšet hlasy Magdalény Korcové a Jaroslava Gavendy. Kontrolu výstupů zajišťuje tým editorů pracujících z domova tak, aby práci mohly vykonávat také osoby na rodičovské dovolené. (Seznam Zprávy, 2024)

Právní smlouvy týkající se využití syntetických hlasů obsahují licence, které mají určité trvání s možností jejich prodloužení. „Vícero speakerů (hlasových talentů, pozn. aut.) bylo v posledním roce prezentováno na produktech, na kterých jsme se dohodli. Smlouva rovněž obsahuje klauzuli o férovém využití, které nezpůsobuje urážky na cti či jinou stigmatizaci autora. Seznam s licencovanými autory hlasů udržuje dobré vztahy a zatím nedošlo k žádnému problému,“ objasňuje Diana Hlaváčová.

Jak dále vysvětluje Špaček, v závislosti na produkci obsahu se denně syntetizuje padesát až sedmdesát článků, z toho přibližně čtyřicet až padesát je syntetizováno v rámci redakce Seznam Zprávy. Ještě před využitím syntetických hlasů byli využíváni lidští předčítači, kteří četli audioverze vydaných textů. „Původně měli editoři klíč, podle něhož denně vybrali čtyři až šest textů, které ve studiu namlouvali lidští předčítači. Systém Stela, který jsem postavil, umožňuje komukoliv s přístupem poměrně jednoduše vytvořit z jakéhokoli článku audio,“ kontruje Špaček. Nyní se primárně syntetizují nejdůležitější úvodní články, až poté texty převzaté či upravené.

Navzdory řadě nástrojů Špaček tvrdí, že automatizovaná žurnalistika v Seznamu kromě načítání článků není. Dodává ale, že rutinní činnosti je možné technologiemi nahradit už dnes. „*Když máte skvělého moderátora, nenahradíte ho. Takhle brzo určitě ne. A především to ani nechcete. V Seznam Zprávách nebylo naším cílem snižovat stavy redaktorů, protože jsme disponovali technologickými hračkami. Záměrem bylo automatizovat rutinní práci novinářů, aby se mohli věnovat své příslušné novinářské práci. AI sice bude prohledávat rejstříky, ale na rozdíl od lidí nemá kontextové vidění světa. Zatím. Aby se tak stalo, musíte takový systém správně postavit. Velmi brzo se to povede. Umělá inteligence nahradí novináře ve smyslu takzvaných ‚překlápěčů četky‘. V příštích letech ale nevidím AI s potenciálem nahradit žurnalisty, kteří někam zavolají, získají informaci a uvedou ji do kontextu.*“ Novináři se tak dle Špačka stanou ve spolupráci s nástroji automatizace silnějšími.

Hlaváčová podotýká, že odbouráváním rutinní práce se Seznam snaží zaměstnancům šetřit čas na činnosti s vyšší přidanou hodnotou. „*Kvalitní lidé si vždy najdou něco, čím mohou firmu posunout dál,*“ doplňuje s tím, že hlas Hacsiko vznikl z důvodu, aby si moderátorka Barbara Hacsí mohla „*odběhnout*“ za novými zkušenostmi v relaci Big Brother. „*Bára se vrátila a stále moderuje, žádný syntetický hlas její důvtip a skvělou povahu, která se propisuje do vysílání, nenahradil,*“ říká Hlaváčová.

Spíše než nahrazení novinářů umělou inteligencí je podle Jiřího Špačka bezprostřednější obava, že tradiční mediální svět bude upozaděn před sociálními médii. „*Politici a firmy se naučili používat sociální média. Myslí si, že média nepotřebují, ale to je chyba. Společnosti hrozí, že u toho upustí, a nakonec se stane loutkou lidí, kteří drží nad sociálními sítěmi moc. Sociální média mají nyní obrovský vliv. Když se však podíváme na to, odkud lidé berou informace, vycházejí buďto z původních médií, nebo jde o příspěvky na sítích, které ale vycházejí z těchto původních zdrojů,*“ zdůrazňuje.

Řečové technologie se neustále vyvíjejí. Podle Špačka již nyní dokáží algoritmizované nástroje z řeči poznat gender, věk, ale také emoci. „*Akademici zabývající se řečovými technologiemi bývají perfekcionista. Mají a zároveň nemají pravdu, že umělá inteligence nedokáže rozpoznat emoce. Lidský hlas považují za náš nejkompexnější komunikační prostředek, je v něm všechno, AI systémy dneska dokáží rozpoznat, zdali je člověk rozčilený či zlomený. Co však technologie nedokáží, je odhalit předstírání emoci na úrovni herců,*“ přibližuje svůj pohled Špaček.

S rychlostí vývoje se podle něj bude rapidně měnit způsob, jakým řečové technologie chápou lidský projev. Budoucnost Špaček vidí v sestrojení komplexních systémů disponujících celkovým pozadím osob. „*Takový robot bude vědět, kdo jsem, jaké je mé pozadí, kde jsem mluvil a co jsem říkal. Informace o mně bude dávat do kontextu mé minulosti, která bude dostupná. A to mluvím pouze o hlase,*“ představuje si budoucnost.

## **Shrnutí**

Technologická společnost Seznam.cz, jejíž součástí jsou mimo jiné zmíněný zpravodajský server Seznam Zprávy i magazínové servery Garáž.cz a Proženy.cz, implementaci řečové syntézy vnímá jako prostředek ke kvalitnější distribuci obsahu a automatizaci rutinních úkolů, nikoli jako technologii s potenciálem nahradit skutečné novináře a moderátory. Na vývoji syntetických hlasů pracuje od roku 2023, přičemž svůj první, moderátorku rádia Expres FM Hacsiko, představila téhož roku. Na něm i na dalších čtyřech hlasech následně Seznam Zprávy spolupracovaly s firmou MAMA AI. Model Hacsiko vytrénován na hlasu moderátorky Barbary Hacsi, pro účely čtení audioverzí textů propůjčili své hlasy také novinářky a novináři Silvie Friedmannová, Magdaléna Korcová, Pavel Vondra a Jaroslav Gavenda. Hlasy čtyř novinářů jsou od prosince 2023 využívány na webech Seznam Zprávy, Proženy.cz a Garáž.cz, přičemž dva z těchto hlasů od března 2024 předčítají zpravodajské relace na rádiu Expres FM a Classic Praha. Zjara roku 2023 se pak vedení Seznamu rozhodlo vyvinout vlastní syntetický hlas nezávislý na outsourcingu.

Bývalý zástupce ředitele pro AI v redakci Seznam Zprávy Jiří Špaček popsal úskalí týkající se vývoje syntetizovaných hlasů. Během nahrávání záznamů hlasových talentů je dle něj věnována pozornost konkrétní intonaci či tempu, přičemž cílem je hlas přizpůsobit konkrétním formátům buďto v rádiu nebo na zpravodajském serveru obsahující komplexní texty. Redakce využívají sadu vlastních nástrojů určených pro řízení a distribuci audioobsahu – Xena, Edna a Stela. Výsledný obsah však podléhá kontrole člověka. Firma podle produktové manažerky Diany Hlaváčové plánuje postupně rozšířit portfolio hlasů a produktů, a řečové technologie v budoucnu implementuje i do jiných sfér mimo obsahovou platformu.

## Shrnutí výzkumu

Cílem této diplomové práce bylo zjistit, jaký přístup uplatňují vybraná česká média vůči technologii syntézy hlasu. Problematika je podstatná z důvodu automatizace odvětví žurnalistiky technologií strojového učení a potenciálu transformovat tvorbu a distribuci mediálního obsahu k cílovému publiku.

Teoretický rámec diplomové práce nastínil počátky ustanovujícího oboru umělé inteligence a prozkoumal jednotlivé perspektivy plynoucí z jejího využití v mediálním odvětví. Exkurz zahrnoval perspektivy automatizované žurnalistiky, její aplikaci v provozu mediálních organizací i limity a výzvy plynoucí z jejího potenciálu.

Pro splnění cíle diplomové práce byla stanovena hlavní výzkumná otázka: „*Jaký postoj vybrané české redakce uplatňují k využívání syntézy hlasu?*“ Diplomová práce se snažila najít odpovědi také na dvě dílčí výzkumné otázky: *Jaké nástroje a služby mediální organizace využívají pro zpracování syntetických hlasů? Jakou roli bude syntéza řeči hrát v budoucím fungování redakcí?*

Praktická část diplomové práce zahrnuje analýzu kvalitativních polostrukturovaných rozhovorů s účastníky výzkumného, produkčního i implementačního procesu v konkrétních mediálních organizacích dohlížející na zavádění technologií umělé inteligence a především řečových technologií, specificky zahrnující syntézu řeči. Jmenovitě šlo o projektovou manažerku Českého rozhlasu Dalii Moudrou, zástupce šéfredaktora Deníku N Lukáše Wernera, současnou produktovou manažerku Seznam.cz Dianu Hlaváčovou a bývalého ředitele pro rozvoj Seznam Zpráv Jiřího Špačka. Pro účely specifického ukotvení tématu syntézy řeči v českém kontextu a jejího vývoje jsem oslovil odborníky na vývoj syntetizovaných hlasů ze soukromé společnosti MAMA AI, která spolupracuje s řadou českých mediálních organizací, a to včetně těch, jimž se věnuje tato diplomová práce. V tomto případě šlo o polostrukturované rozhovory vedené s Josefem Vopičkou a Jiřím Vrobelem.

Respondenti z firmy MAMA AI popsali princip fungování technologií syntézy řeči (text-to-speech) i rozpoznávání řeči (speech-to-text). Vylepšovat kvalitu syntetických hlasů je dle nich možné v rovině kvality hlasu, výslovnosti neznámých slov a zdokonalování konkrétních emocí. Technologie jsou v rámci mediálních organizací uplatňovány v nástrojích transkripce, tvorby syntetických hlasů využívaných pro tvorbu audioobsahu i audioverzí článků. Své služby v České

republice MAMA AI poskytuje například médiím Seznam Zprávy, Český rozhlas, Česká televize a TV Prima. Společnost je schopna syntetizovat hlasy z přibližně pětihodinových audionahrávek pořízených v profesionálním studiu, trénink modelu s využitím neuronových sítí trvá zhruba tři týdny. Právně je využití cizího hlasu ze strany MAMA AI striktně ošetřeno a podléhá etickým účelům. Respondenti však nabádají k opatrnosti, neboť přílišné využití syntetických asistentů například prohlubuje pocit osamění u osob, kterým asistenti mají pomoci proti osamělosti či introverzi.

Pro Deník N je text-to-speech součástí budoucího rozvoje, technologie umožňuje kvalitnější zpřístupnění svého rozsáhlého textového obsahu pro cílové publikum. Zpočátku server pracoval s lidskými předčítači, kvůli řadě problémů zahrnující časově náročný dohled editorů, chyby ve výslovnosti, ale i vysoké finanční náklady přešel k robotickému hlasu. Redakce až do letošního roku využívala nástroj Azure od společnosti Microsoft. Protože se navzdory nízkým nákladům potýkala s technickými omezeními a problémy s výslovností, rozhodla se přejít ke společnosti ElevenLabs, jež nabízí pokročilejší hlasové modely, multilingvální podporu a možnost syntetizovat vlastní hlasové talenty. Deník N proto syntetizuje a následně zařadí do provozu hlasy čtyř členů redakce - konkrétně hlasy Lenky Vrtiškové Nejezchlebové, Jany Ciglerové, Pavla Tomáška a Michaela Švece. K výsledným audioverzím článků Deník N přidá upozornění, že hlasy nejsou reálné. Se svými hlasovými talenty redakce uzavře jednoduché smlouvy, v případě externí spolupráce mimo mediální organizaci. Z pozice menší mediální organizace s omezeným rozpočtem si tedy Deník N může dovolit pouze částečný in-house vývoj vlastních technologických novinek, mimo jiné v závislosti na vývojářských kapacitách sesterské redakce slovenského Denníku N.

Český rozhlas zavedl syntézu řeči v květnu 2024, využití technologie našla při tvorbě audioverzí článků, a mediální organizace vidí technologii syntézy řeči jako doplňkový nástroj s potenciálem rozšířit možnosti audio obsahu, nikoliv nahrazovat lidské moderátory. Do provozu byl prvně nasazen hlas moderátora Pavla Prouzy, roku 2025 také ženský hlas, u nějž probíhá optimalizace. V případě redakce iRozhlas, kde se hlasy pro tyto účely využívají, proces výběru článků k syntéze probíhá zatím spíše nahodile, bez jasně daných pravidel. Redakce však v tomto ohledu směřuje ke konkrétnějšímu rámci. Smlouvy, které Český rozhlas se svými hlasovými talenty uzavírá, obsahují licence o propůjčení hlasu, dobu, po kterou je možné hlas využívat, a finanční plnění ve vztahu k hlasovému talentu. Na zpracování dvou syntetických hlasů Český rozhlas spolupracoval s firmou

SpeechTech. V rámci grantového projektu pak Rozhlas spolupracuje i s firmou MAMA AI, která vytvořila další ženský hlas, určený pro výzkum s cílem vyvinout audio asistenta, který uživatelům Českého rozhlasu umožní větší míru interakce a personalizace obsahu. Celková implementace syntézy řeči se opírá o přísné vnitřní směrnice pro užití umělé inteligence, aby byl zachován závazek k udržení kvality audio obsahu. Budoucí směřování Český rozhlas zaměří na rozšíření možnosti interakce se svým obsahem pomocí hlasových asistentů.

Seznam.cz vnímá implementaci syntetických hlasů coby prostředek pro kvalitní distribuci obsahu a automatizaci rutinních úkolů, a to bez potenciálu nahradit skutečné novináře a moderátory. Vývoj umělých hlasů zde probíhá od roku 2023, ve spolupráci s firmou MAMA AI bylo natrénováno celkem pět hlasů využívaných v rádiích Expres FM a Classic Praha a pro čtení audioverzí článků na webech Seznam Zprávy, Proženy.cz a Garáž.cz. Své hlasy poskytli Barbara Hacsí, Silvie Friedmannová, Magdaléna Korcová, Pavel Vondra a Jaroslav Gavenda. Redakce využívají sadu vlastních nástrojů určených pro řízení a distribuci audioobsahu – Xena, Edna a Stela. Veškerý generovaný obsah podléhá kontrole člověka. Smlouvy, které mediální organizace s hlasovými talenty uzavřela, obsahují licence s určitým trváním s možností prodloužení a klauzuli o férovém užití. Relace hlasu Hacsíko obsahují upozornění na syntetický původ, jiné, jako jsou zpravodajské relace hlasů Magdalény Korcové a Jaroslava Gavendy či audioverze článků na zmíněných webech z důvodů původních tiskových zpráv a článků upozornění neobsahují. V roce 2024 se Seznam.cz následně rozhodl vyvinout vlastní řešení syntetického hlasu nezávislého na outsourcingu služeb. Respondenti popsali specifika vývoje syntetizovaných hlasů a budoucí vývoj zahrnující rozšíření portfolia vlastních hlasů i do jiných sfér mimo obsahovou platformu.

Konvergence přístupu zkoumaných subjektů stojí na několika úrovních. Všechny tři mediální organizace vnímají technologii text-to-speech jako nástroj pro zvýšení dostupnosti svého obsahu, zejména pro uživatele preferující audio formáty. Redakce Deníku N a Seznamu prošly fází využívání lidských předčítačů, poté se vzhledem k časově náročné údržbě tohoto modelu přesunuli k levnějšímu řešení v podobě robotických hlasů. Všechny tři mediální organizace začaly investovat do vývoje vlastních syntetických hlasů, možnosti vyvíjení se však vzhledem k odlišným způsobům financování vývoje liší. Přístup zkoumaných organizací se sbíhá i na outsourcingu, tedy vyčlenění konkrétní činnosti na jinou firmu. V tomto případě se všechny subjekty v určité fázi spoléhaly či stále spoléhají na externí společnosti, specializující se na řečové technologie. Organizace zároveň investují do vytváření syntetických hlasů založených na hlasech konkrétních osob spojených

s jejich redakcemi a jsou si vědomy limitů současné technologie. Přínosy text-to-speech však považují za dostatečné pro nasazení do provozu. Žádný ze zmíněných nástrojů není plně autonomní, proces tvorby a kontrola výsledků vždy vyžadují určitý editorský dohled. Všechny tři mediální organizace deklarují, že cílem využití technologie syntézy řeči není nahradit základní novinářskou či moderátorskou činnost, ale spíše doplnit nebo automatizovat žurnalistickou rutinní práci.

V určitých směrech jsou však organizace divergentní. Seznam.cz jako technologická firma směřuje k plně internímu vývoji syntézy hlasu. ČRo naopak vývoj outsourcuje a soustředí se na audio expertizu. Deník N využívá externí dodavatele a částečně interní kapacity vývoje, avšak nebuduje fundament technologie syntézy řeči. Seznam.cz implementuje technologii TTS masivně a inovativně, zato Český rozhlas je konzervativní a využívá syntézu hlasu pouze jako doplněk v omezeném rozsahu. Deník N stojí mezi těmito médii, snaží se o plné pokrytí článků, je však limitován technologií a zdroji. V případě Seznamu je vedle přístupnosti pro uživatele patrný i komerční motiv v podobě reklamního prostoru. U ČRo dominuje étos veřejnoprávní služby. Pro Deník N je motivem dostupnost psaného obsahu pro své předplatitele. Označování syntetického obsahu se napříč organizacemi liší. Zatímco Deník N a Český rozhlas již explicitně označují či plánují upozornění optimalizovat, u Seznamu je syntetický obsah označován nejednotně a částečně.

Odborné prameny a literatura, z které čerpala tato diplomová práce, se až na výjimky věnovala zahraničnímu a obecnému kontextu dopadů umělé inteligence na žurnalistiku a profesi novinářů. Z provedeného výzkumu však lze uvedenou teorii uplatnit i v diskurzu českých médií. Implementace v mnoha doménách žurnalistiky skutečně odpovídá perspektivám autorů, jako jsou Deuze, Beckett, Marconi či Graefe. Přestože tato diplomová práce není konkrétně zaměřena na pohled novinářů vůči nástrojům umělé inteligence, z dlouhodobého pozorování fenoménu vnímám mezi novináři obezřetnost, obavy, ale také otevřenost vůči novým technologiím.

Tato diplomová práce odpovídá nejen na definované cíle a výzkumné otázky, poskytuje i vhled do jednotlivých perspektiv užívání syntézy hlasu a řečových technologií ve vybraných mediálních organizacích, zahrnující veřejnoprávní i soukromá média. Poskytuje exkurz do začátků implementace technologie text-to-speech i speech-to-text ve vybraných českých médiích, ale také nabízí vhled do limitů řečových technologií, zpracování smluv mezi redakcemi a hlasovými talenty, dotýká se etického rámce potenciálního nahrazení novinářů nástroji umělé inteligence, řeší i upozornění na umělé hlasy a s tím spojenou důvěru posluchačů syntetizovaného obsahu,

zpracování řeči pro konkrétní mediální formáty zasazuje do kontextu českého jazyka, navíc doplněného o perspektivu účastníků výzkumného a produkčního procesu, kteří objasnili i fundamentální fungování řešených řečových technologií. Všechny tyto detaily poskytují příležitost dalším badatelům zkoumat dílčí aspekty řečových technologií v mediální oblasti.

## Závěr

Umělá inteligence bude mít na mediální odvětví disruptivní dopady. Co se ještě v minulé dekádě zdálo jako mlhavý opar daleké budoucnosti, dnes se díky strojovému učení transformuje v reálné důsledky implementace umělé inteligence. Sám jsem jakožto novinář ještě před třemi lety přistupoval k automatizované žurnalistice se značnou rezervou. Dnes díky pokročilým nástrojům vidím jejich potenciál šetřit hodiny času, ale také schopnost toto povolání zcela proměnit.

Z dlouhodobého pozorování automatizace žurnalistiky, a zvláště intenzivně v posledním roce v rámci zpracování této diplomové práce, pozoruji propojení teoretické perspektivy akademiků se skutečným stavem věci. Studie zabývající se dopady automatizace na práci novinářů již v minulé dekádě zmiňují, že umělá inteligence převezme rutinní práci a zvýší efektivitu tohoto povolání. Realizace výzkumu potvrzuje trend automatizace žurnalistiky. Průběh algoritmizace urychlila i demokratizace generativních nástrojů umělé inteligence, které svými schopnostmi dosavadní automatizaci narušili. Rovněž mě zaujalo, že někteří autoři píší o technologii převodu řeči na text (*speech-to-text*) téměř nekriticky. Praxe přitom ukazuje, že dostupné transkripty sice novinářům ušetří čas, stále však velice často chybují a je nutné přepsané záznamy kontrolovat.

Do širšího tématu však zasahují – zcela oprávněné – obavy novinářů o ztrátu vlastní profesní autonomie. Žurnalistika se v budoucnu promění. Pracovní postupy novinářů budou jiné, možná se změní k nepoznání. Ne hned, nýbrž v dlouhodobém horizontu. V jakémkoliv smyslu dopadů však bude důležité zachovat principy kritické žurnalistiky, nepropadat trendům automatizace jen za cenu ušetření nákladů ze strany jednotlivých titulů.

Jakkoli je předpověď budoucnosti lákavá a zároveň nemožná, domnívám se, že ze stávajících trendů lze vytyčit směry, kterými se žurnalistika vydá. Dovolím si proto nastínit možné scénáře, jak umělá inteligence promění práci novinářů – a zpravidla budou tahat za kratší konec. Protože má žurnalistika i mnoho domén, dovolím si následující scénáře ilustrovat na příkladu zpracování textu a zvuku, jimž se věnovala i tato diplomová práce.

Představme si horizont deseti let, tedy roku 2035. Výpočetní výkon datových serverů bude již natolik vysoký a dostupný, že i relativně malé tituly si budou moci dovolit provoz vlastních neuronových sítí a vytrénovat si vlastní nástroje pro zpracování textu i zvuku, které budou otiskem daného titulu, jednotlivých autorů a jejich charakteristického stylu projevu – hlasového i psaného.

Redakční systémy tuzemských médií totiž disponují datasety tisíců a desetitisíců článků i zvukového materiálu, a za deset let se jejich počet znásobí. Členové redakcí mohou být nadevše konzervativní, technologický vývoj je ale dožene. Stroj začne být seriózní konkurencí pro každého redaktora i moderátora. Až se začne smazávat rozdíl mezi výkonem lidského novináře a toho umělého, vstoupí do průběhu vývoje tři zásadní proměnné – role vedoucích manažerů titulů, společnosti a novinářovy „kreativity“.

Soukromé mediální domy fungují, pochopitelně jako každá jiná firma v kapitalistickém světě, na modelu růstu. Jakkoli zaznívá kritika, že tento model není s kvalitní novinářskou prací slučitelný, vedoucí management mediálních domů tlačí na nekonečně vzrůstající návštěvnost a zisky. (Hrdina, 2023) Pokud uvidí, že nástroje automatizace tento růst zachovají (tzn. čtenáři si na automatizovaný obsah navyknu), šli by proti svým zájmům, kdyby do provozu redakcí takové nástroje neimplementovali. Vše se bude řídit tradičními tržními principy. Pokud k automatizaci žurnalistiky vedoucí manažeři přistoupí netransparentně a nebudou s řadovými pracovníky redakcí o těchto změnách komunikovat, hrozí, že se redakce propadnou do hluboké frustrace.

Druhou roli zastane přístup společnosti k automatizované žurnalistice. Starší lidé jsou podle průzkumu z roku 2024 zdrženlivější vůči využití AI v žurnalistice, mladší lidé (do 44 let) vidí její využití pozitivněji. Bezmála tři čtvrtiny Čechů však preferují práci lidských novinářů před umělou inteligencí. (Moravec, 2024) Občanská společnost může vytvářet tlak na mediální domy, aby byl zachován stav, kdy obsah tvoří výhradně lidé (kteří ale i v případě automatizace musí výstupy kontrolovat). Aby byly nástroje automatizace kvalitní, neuronové sítě musí zpracovat rozsáhlé množství dat. V oblasti syntézy řeči jsem ještě nedávno zaznamenal názory, že tvorba kvalitních nástrojů pro český kontext je téměř nemožná, protože čeština je malý a nevýznamný jazyk (a platí to i v případě trénování textových chatbotů). Jak ukazují studie, modely TTS jsou skutečně těžce škálovatelné pro všechny jazyky, neboť pozornost se zřídka věnuje malým jazykům. (Amalas et. al., 2024) Jak lze ale v současné době vidět, s rozeznáním syntetických hlasů od skutečných má problém mnoho konzumentů mediálního obsahu. (Kasík, 2023) Domnívám se, že s větším prolínáním nových technologií do sfér běžného života dojde – minimálně mezi mladší generací – k jejich přijetí i na úrovni konzumace novinářského obsahu.

Celá problematika není jednoduchá a bude záležet i na dalších faktorech – jinak budou automatizované nástroje přijímat veřejnoprávní média, jinak soukromá, jejichž model je někdy navíc založen na platební bráně. To ostatně plyne i z provedené výzkumné části této práce.

Publikum takových titulů bude k zavádění automatizace konzervativnější a bude vyžadovat práci lidských novinářů. Zároveň si čtenáři spojují text s konkrétními autory. Je také možné, že za deset let bude výhradně lidská doména součástí identity konkrétních titulů, na „lidské novinářině“ si média založí svůj brand.

Limity automatizace však vidím i ve ztrátě pracovního smyslu novinářů. AI nástroje jednoznačně urychlí a zefektivní jejich práci, denně tak napíší násobně víc textů. Rizika zde vnímám na dvou úrovních – mediální domy sice vyprodukují násobek textů, jenomže lidská pozornost je omezená. Média nebudou produkovat více času pro jejich čtení. Zároveň však zvyšování produktivity může narazit na stav, kdy novináři za den „projde pod rukama“ množství textů, s kterými kvůli zrychlenému modelu stráví jen zlomek času, a bude otázkou, zdali jejich téma dokáže vstřebat a kriticky nad nimi uvažovat v širším kontextu. Novinářův pracovní postup se může zredukovat výhradně na psaní promptů a příkazů do redakční AI. V takovém případě buď dojde k novinářově vyhoření, nebo v lepším případě ke změně dosavadního přístupu k práci.

Novináři se proti automatizaci mohou vymezit vlastním kreativním myšlením, umělá inteligence však v budoucnu odhalí rozdíl mezi zdánlivou a skutečnou kreativitou a mezi zdánlivou a skutečnou žurnalistikou. Ty skutečné proměnné zatím nenahradí. Novinářské texty jsou nicméně v základu žánrově konvenční, a AI se tyto konvence může naučit. AI systémy v budoucnu dokážou vyhledat témata, sepsat okruhy otázek k rozhovorům a přepsat je, reagovat na odpovědi, napsat z předložených dat články a distribuovat je čtenářům. Vývojáři nástrojů automatizace sice mezitím tvrdí, že chtějí ušetřit novinářům práci, zároveň však vytvářejí nástroje, které jejich (rutinní) práci nahrazují. Aby novináře algoritmy neporazily, musí své profesi přinášet vyšší hodnotu. Umělá inteligence však ještě dlouho bude narážet na schopnost kontextového vidění světa. Podoba budoucí profese novinářů se bude muset vymykat automatizovatelnému standardu. Bude jiná. Tváří v tvář obavám však bude přijímání umělé inteligence náročné, leč nevyhnutelné. Věřím, že mnoha skutečným novinářům se to podaří, neboť budou nadále schopni odhalovat skrytou podstatu světa tam, kde to nedokáže umělá inteligence.

## Zdroje

### Literatura

HENDL, Jan, 2005. *Kvalitativní výzkum: základní metody a aplikace*. Praha: Portál. ISBN 80-736-7040-2.

MARCONI, Francesco, 2020. Online. In: *Newsmakers: artificial intelligence and the future of journalism*. New York: Columbia University Press. ISBN 978-0-231-19137-1. [cit. 2025-05-09].

MINSKY, Marvin Lee a PAPERT, Seymour, c1969. *Perceptrons: an introduction to computational geometry*. Expanded ed. Expanded Edition. Cambridge, Mass.: MIT Press. ISBN 978-0-262-63111-2.

MORAVEC, Václav, 2020. *Proměny novinářské etiky*. Společnost. Praha: Academia. ISBN 978-80-200-3111-2.

SEDLÁKOVÁ, Renata. *Výzkumy médií: nejužívanější metody a techniky*. Praha: Grada, 2014. ISBN 978-80-247-3268-9.

ŠVAŘÍČEK, Roman a ŠEĐOVÁ, Klára, 2007. *Kvalitativní výzkum v pedagogických vědách*. Praha: Portál. ISBN 978-80-7367-313-0.

### Elektronické zdroje

AMALAS, Asma; GHOGHO, Mounir; CHETOUANI, Mohamed a OULAD HAJ THAMI, Rachid, 2024. A multilingual training strategy for low resource Text to Speech. Online. Dostupné z: <https://doi.org/10.48550/arXiv.2409.01217>. [cit. 2025-05-09].

Artificial Intelligence in Journalism, 2025. Online. In: Center for News, Technology & Innovation. Dostupné z: <https://innovating.news/article/ai-in-journalism/>. [cit. 2025-05-05].

BBC, 2015. ALTO - a multilingual journalism tool. Online. BBC. Dostupné z: <https://www.bbc.co.uk/rdnewslabs/projects/alto/>. [cit. 2025-05-08].

BBC(b), 2020. BBC Global News launches AI-powered synthetic voice which ‘reads’ articles on BBC.com. Online. BBC. Dostupné z: <https://www.bbc.co.uk/mediacentre/worldnews/2020/life-project>. [cit. 2025-05-08].

BECKETT, Charlie a YASEEN, Mira, 2023. Generating Change: A global survey of what news organisations are doing with. Online. JournalismAI. Dostupné z: <https://www.journalismai.info/research/2023-generating-change>. [cit. 2025-05-05].

BUMBÁLKOVÁ, Blanka, 2023. Projekt Gott navždy je nejposlouchanějším dílem v historii Českého rozhlasu. Online. Český rozhlas. Dostupné z: <https://informace.rozhlas.cz/projekt-gott-navzdy-je-nejposlouchanejsim-dilem-v-historii-ceskeho-rozhlasu-9063557>. [cit. 2025-05-08].

CAMBRE, Julia; COLNAGO, Jessica; MADDOCK, Jim; TSAI, Janice a KAYE, Jofish, 2020. Choice of Voices: A Large-Scale Evaluation of Text-to-Speech Voice Quality for Long-Form Content. Online. CHI Conference on Human Factors in Computing Systems (CHI '20). Dostupné z: <https://doi.org/10.1145/3313831.337678>. [cit. 2025-05-08].

CARLSON, Matt, 2015. The Robotic Reporter. Online. Digital Journalism. S. 416-431. Dostupné z: <https://doi.org/10.1080/21670811.2014.976412>. [cit. 2025-05-09].

CARTER ROSS, T., 2024. BBC Sounds Begins AI-Powered Subtitles Trial. Online. Radio World. Dostupné z: <https://www.radioworld.com/tech-and-gear/products/bbc-sounds-begins-ai-powered-subtitles-trial>. [cit. 2025-05-06].

COOK, Clare; GARCIA, Emiliana; GYULNAZARYAN, Heghine; MELANO, Juan a PARUSINSKI, Jakub, 2021. The next wave of disruption: Emerging market media use of artificial intelligence and machine learning. Online. International Media Support. Dostupné z: <https://www.mediasupport.org/publication/the-next-wave-of-disruption-emerging-market-media-use-of-artificial-intelligence-and-machine-learning/>. [cit. 2025-05-01].

CYBERNETICZOO, 2009. 1951 – SNARC Maze Solver – Minsky / Edmonds (American). Online. Cyberneticzoo. Dostupné z: <https://cyberneticzoo.com/mazesolvers/1951-maze-solver-minsky-edmonds-american/>. [cit. 2025-05-09].

ČESKÝ ROZHLAS, 2023. Na hranici technologických možností. Jak vznikl digitální otisk hlasu Karla Gotta? Online. Český rozhlas. Dostupné z: <https://dvojka.rozhlas.cz/na-hranici-technologickych-moznosti-jak-vznikl-digitalni-otisk-hlasu-karla-gotta-9007323>. [cit. 2025-05-08].

ČESKÝ ROZHLAS, 2024. Pravidla pro užívání umělé inteligence v Českém rozhlasu. Online. Český rozhlas. Dostupné z: <https://informace.rozhlas.cz/pravidla-pro-uzivani-umele-inteligence-v-ceskem-rozhlasu-9169447>. [cit. 2025-05-08].

ČTK, 2023. ČTK připravila regule k AI, prvky automatizace jsou ve zpravodajství už 30 let. Online. In: Česká tisková kancelář. Dostupné z: <https://www.ctk.cz/novinky/?id=3587>. [cit. 2025-05-01].

DENNIS, Michael Aaron, 2024. Marvin Minsky. Online. In: Britannica. Dostupné z: <https://www.britannica.com/biography/Marvin-Minsky>. [cit. 2025-05-01].

DEUZE, Mark a BECKETT, Charlie, 2022. Imagination, algorithms and news: developing AI literacy for journalism. Online. Digital Journalism. Roč. 10. Dostupné z: <https://doi.org/10.1080/21670811.2022.2119152>. [cit. 2025-05-01].

FEBRI SONNI, Alem; HAFIED, Hasdiyanto; IRWANTO, Irwanto a LATUHERU, Rido, 2024. Digital Newsroom Transformation: A Systematic Review of the Impact of Artificial Intelligence on Journalistic Practices, News Narratives, and Ethical Challenges. Online. Journalism and Media. Roč. 5. Dostupné z: <https://doi.org/10.3390/journalmedia5040097>. [cit. 2025-05-05].

GÉLA, František a MORAVEC, Václav, 2024. Umělá inteligence ve službách Českého rozhlasu. Film. Online. In: Český rozhlas. Dostupné z: <https://www.irozhlas.cz/signal-a-sum-dokument>. [cit. 2025-05-08].

GRAEFE, Andreas, 2016. Guide to Automated Journalism. Online. Tow Center for Digital Journalism, Columbia Journalism School. Dostupné z: [https://www.cjr.org/tow\\_center\\_reports/guide\\_to\\_automated\\_journalism.php](https://www.cjr.org/tow_center_reports/guide_to_automated_journalism.php). [cit. 2025-05-01].

GUTIÉRREZ-CANEDA, Beatriz; LÓPEZ-GARCÍA, Xosé a VÁZQUEZ-HERRERO, Jorge. AI application in journalism: ChatGPT and the uses and risks of an emergent technology. Online. Profesional de la información. Roč. 32. Dostupné z: <https://doi.org/10.3145/epi.2023.sep.14>. [cit. 2025-05-05].

HEISELBERG, Lene; BLOM, Jonas a DALEN, Arjen van, 2022. Automated News Reading in the Neural Age: Audience Reception and Perceived Credibility of a News Broadcast Read By a Neural Voice. Online. Journalism Studies. Roč. 23, s. 896-914. Dostupné z: <https://doi.org/10.1080/1461670X.2022.2052346>. [cit. 2025-05-08].

HRDINA, Matouš, 2023. Pod čarou: Facebook dal a zase vzal. Jak digitální média přišla na buben. Online. Seznam Zprávy. Dostupné z: <https://www.seznamzpravy.cz/clanek/tech-pod-carou-facebook-dal-a-zase-vzal-jak-digitalni-media-prisla-na-hul-230895>. [cit. 2025-05-01].

JAVŮREK, Karel, 2023. Lidé si k AI vytvořili citový vztah. Chtěli, aby jim umělý Tomáš volal denně. Online. Wired. Dostupné z: <https://www.wired.cz/clanky/lide-si-k-ai-vytvorili-citovy-vztah-chteli-aby-jim-umely-tomas-volal-denne>. [cit. 2025-05-09].

JONG WOOK, Kim, 2024. GitHub - openai/whisper: Robust Speech Recognition via Large-Scale Weak Supervision. Online. GitHub. Dostupné z: <https://github.com/openai/whisper>. [cit. 2025-05-07].

KASÍK, Pavel, 2023. Poznejte robota: Seznam Zprávy testují v rádiu stroj, který mluví jako člověk. Online. Seznam Zprávy. Dostupné z: <https://www.seznamzpravy.cz/clanek/tech-ai-umela-inteligence-poznejte-robota-seznam-zpravy-testuji-v-radiu-stroj-ktery-mluvi-jako-clovek-232868>. [cit. 2025-05-08].

KOENECKE, Allison; SEO GYEONG CHOI, Anna; X. MEI, Katelyn; SCHELLMANN, Hilke a SLOANE, Mona, 2024. Careless Whisper: Speech-to-Text Hallucination Harms. Online. The 2024 ACM Conference on Fairness, Accountability, and Transparency. Dostupné z: <https://doi.org/10.48550/arXiv.2402.08021>. [cit. 2025-05-07].

KOTENIDIS, Efthimis a VEGLIS, Andreas, 2021. Algorithmic Journalism—Current Applications and Future Perspectives. Online. Journalism and Media. S. 244-257. Dostupné z: <https://doi.org/10.3390/journalmedia2020014>. [cit. 2025-05-01].

KOUBSKÝ, Petr, 2019. Umělá inteligence vznikla v polovině dvacátého století, když se ještě věřilo na ústřední mozek lidstva. Online. Deník N. Dostupné z: <https://denikn.cz/63598/umela-inteligence-vznikla-v-polovine-dvacateho-stoleti-kdyz-se-jeste-verilo-na-ustredni-mozek-lidstva/>. [cit. 2025-05-01].

KUČEROVÁ, Daniela, 2024. Pust'íte si články Seznam Zpráv do uší. Naši roboti čtou za vás. Online. Seznam Zprávy. Dostupné z: <https://www.seznamzpravy.cz/clanek/redakce-pustte-si-clanky-seznam-zprav-do-usi-nasi-roboti-ctou-za-vas-244688>. [cit. 2025-05-08].

LÁB, Filip, 2019. Umělá inteligence v žurnalistice. Online. In: Mediální vzdělání a žurnalistika umělé inteligence. Dostupné z: [aijournalism.fsv.cuni.cz](http://aijournalism.fsv.cuni.cz). [cit. 2025-05-01].

MAGEE, Tamlin, 2024. How BBC Sounds is using GenAI to boost accessibility in audio. Online. Racounter. Dostupné z: <https://www.raconteur.net/technology/how-bbc-sounds-is-using-genai-to-boost-accessibility-in-audio>. [cit. 2025-05-06].

MARCONI, Francesco a SIEGMAN, Alex, 2017. A day in the life of a journalist in 2027: Reporting meets AI. Online. Columbia Journalism Review. Dostupné z: <https://www.cjr.org/innovations/artificial-intelligence-journalism.php>. [cit. 2025-05-05].

MORAVEC, Václav; MACKOVÁ, Veronika; SIDO, Jakub a EKŠTEIN, Kamil, 2020. The Robotic Reporter in The Czech News Agency: Automated Journalism and Augmentation in the Newsroom. Online. Communication Today. Roč. 11, č. 2. Dostupné z: <https://communicationtoday.sk/the-robotic-reporter-in-the-czech-news-agency-automated-journalism-and-augmentation-in-the-newsroom/>. [cit. 2025-05-09].

MORAVEC, Václav, 2024. Ve stínu ChatGPT: proměny vnímání a využívání umělé inteligence českou populací a žurnalistikou. Online. Ipsos. Dostupné z: <https://www.ipsos.com/cs-cz/cesi-umela-inteligence-roste-vyuzivani-obezretny-postoj-pretrvava>. [cit. 2025-05-08].

MÜLLER, Luděk, 2014. Automatické titulkování živých pořadů České televize – současný stav a výhled do budoucna. Online. Helpnet. Dostupné z: <https://www.helpnet.cz/aktualne/ludek-muller-automaticke-titulkovani-zivych-poradu-ceske-televize-soucasny-stav-vyhled-do-0>. [cit. 2025-05-08].

NEWMAN, Nic, 2018. The Future of Voice and the Implications for News. Online. Reuters Institute for the Study of Journalism. Licence: <https://creativecommons.org/licenses/by/2.0/>. Dostupné z: <https://reutersinstitute.politics.ox.ac.uk/news/new-report-future-voice-and-implications-news>. [cit. 2025-05-05].

NEWMAN, Nic a CHERUBINI, Federica, 2025. Journalism, media, and technology trends and predictions 2025. Online. DIGITAL NEWS PROJECT. Dostupné z: <https://reutersinstitute.politics.ox.ac.uk/journalism-media-and-technology-trends-and-predictions-2025>. [cit. 2025-05-05].

NOAIN-SÁNCHEZ, Amaya, 2022. Addressing the Impact of Artificial Intelligence on Journalism: the perception of experts, journalists and academics. Online. Communication & Society. Roč. 35, s. 105-121. ISSN 0214-0039. Dostupné z: <https://doi.org/10.15581/003.35.3.105-121>. [cit. 2025-05-01].

OXFORD ENGLISH DICTIONARY, 2025. Artificial intelligence. Online. Dostupné z: [https://www.oed.com/dictionary/artificial-intelligence\\_n?tl=true](https://www.oed.com/dictionary/artificial-intelligence_n?tl=true). [cit. 2025-05-01].

OXFORD ENGLISH DICTIONARY, 2025. Intelligence. Online. Dostupné z: <https://www.oxfordlearnersdictionaries.com/definition/english/intelligence>. [cit. 2025-05-09].

PAVLIK, John, 2000. The Impact of Technology on Journalism. Online. Journalism Studies. S. 229-237. Dostupné z: <https://doi.org/10.1080/14616700050028226>. [cit. 2025-05-01].

PHELAN, David, 2024. Plaud Note Review: Is This The Best, Slickest, Voice Recorder You Can Buy? Online. Forbes. Dostupné z: <https://www.forbes.com/sites/davidphelan/2024/07/14/plaud-note-review-is-this-the-best-slickest-voice-recorder-you-can-buy/>. [cit. 2025-05-05].

Podcast Technicky vzato: Jak poznat, že nám volá naše babička, a ne AI deepfake?, 2024. Online. In: VUT.cz. Dostupné z: <https://www.vut.cz/vut/f19528/d265620>. [cit. 2025-05-05].

RING OLSEN, Gunhild, 2023. Enthusiasm and Alienation: How Implementing Automated Journalism Affects the Work Meaningfulness of Three Newsroom Groups. Online. Journalism Practice. Roč. 19. Dostupné z: <https://doi.org/10.1080/17512786.2023.2190149>. [cit. 2025-05-05].

SEZNAM ZPRÁVY, 2024. Seznam Zprávy můžete i poslouchat. Až 40 zpráv denně. Online. Seznam Zprávy. Dostupné z: <https://www.seznamzpravy.cz/clanek/redakce-seznam-zpravy-muzete-i-poslouchat-az-40-zprav-denne-247309>. [cit. 2025-05-08].

THOMSON REUTERS. Is it news or noise? Online. Dostupné z: <https://archive.annual-report.thomsonreuters.com/2016/is-it-news-or-noise.html>. [cit. 2025-05-01].

TOMKOVÁ, Šimona, 2025. Text-To-Speech Technologies in Online Media. Online. Media & Marketing Identity. S. 699-707. Dostupné z: <https://doi.org/10.34135/mmidentity-2024-69>. [cit. 2025-05-08].

VAN DALEN, Arjen, 2024. Revisiting the Algorithms Behind the Headlines. How Journalists Respond to Professional Competition of Generative AI. Online. Journalism Practice. S. 1-18. Dostupné z: <https://doi.org/10.1080/17512786.2024.2389209>. [cit. 2025-05-01].

WASHPOSTPR, 2020. The Washington Post to debut AI-powered audio updates for 2020 election results. Online. The Washington Post. Dostupné z: <https://www.washingtonpost.com/pr/2020/10/13/washington-post-debut-ai-powered-audio-updates-2020-election-results/>. [cit. 2025-05-07].

WASHPOSTPR, 2021. The Washington Post integrates Amazon Polly, allowing readers to listen to articles across Post platforms. Online. The Washington Post. Dostupné z: <https://www.washingtonpost.com/pr/2021/05/03/washington-post-integrates-amazon-polly-allowing-readers-listen-articles-across-post-platforms/>. [cit. 2025-05-07].