Jakub Sedláček

Studia nových médií FF UK

Generativní AI pro fotografie, hlas a video.

(aneb pokroky AI v oblasti generování „deepfakes“)

Fotografie

Midjourney (V5.2)

vs

DALL·E (3)

vs

Adobe Firefly (2)

vs

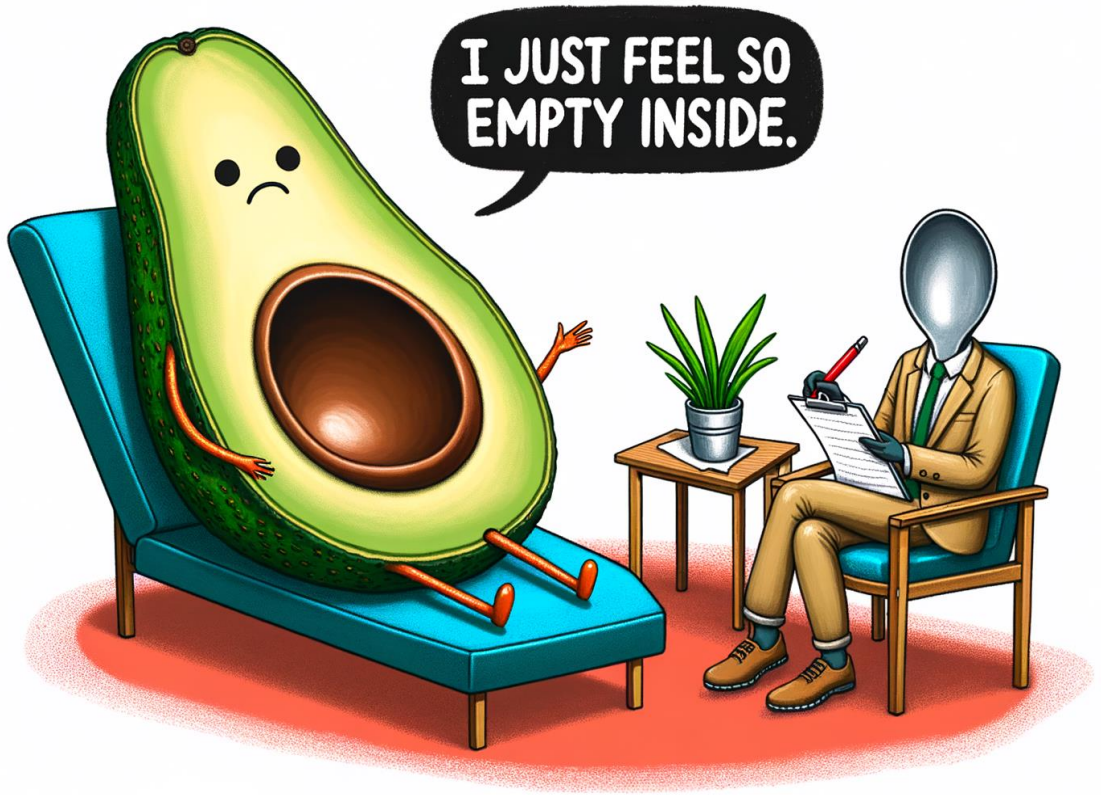
Stable Diffusion (XL)

Autor: Pablo Xavier (3/2023)
Nástroj: Midjourney V5 (?)

Midjourney

- Projekt malého týmu postavený na existujících technologiích, který má ale cutting-edge výsledky
- umí uměleckou krásu spíš než fotky
- placená (od $10/měsíc)
- (zatím) dostupná jen přes Discord

Autor: Jason Matthew Allen (9/2022)
Nástroj: Midjourney V3 (?)



Prompt: An illustration of an avocado sitting in a therapist's chair, saying 'I just feel so empty inside' with a pit-sized hole in its center. The therapist, a spoon, scribbles notes (Autor: OpenAI)

DALL·E 3

- umí kompozičně velmi složité scény a dobře rozumí promptům
- moc neumí fotky
- zdarma v Bingu 15 promptů/den, poté s delším čekáním
- *silná* cenzura

Adobe Firefly 2

- Umí fotky (!)
- Integrace ve Photoshopu (generative fill)
- Na webu zdarma 25 promptů/měsíc
- 100% legální dataset

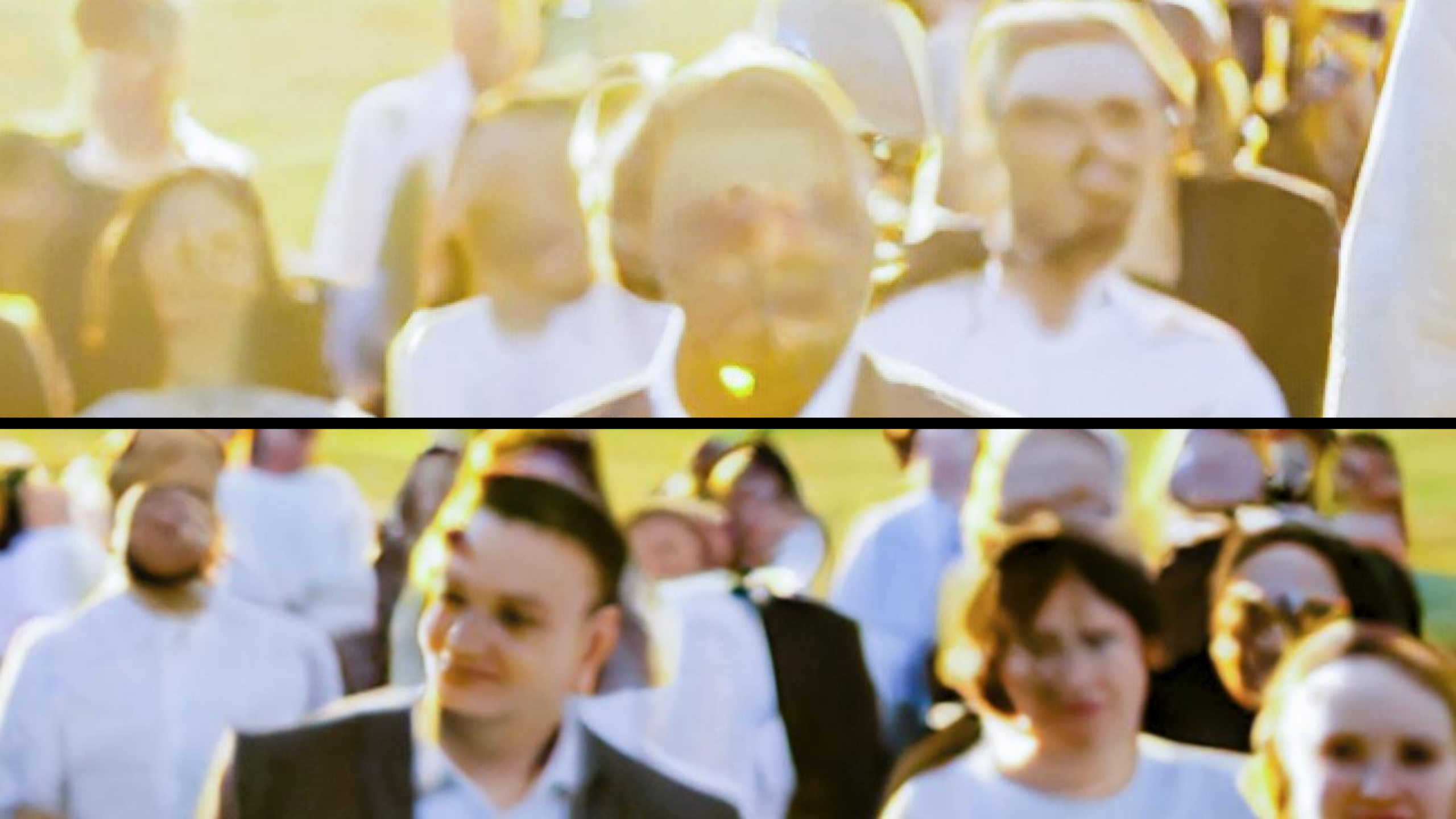
Autor: Licovoda (10/2023)
Nástroj: Stable Diffusion 1.5 (mix modelů epiCPhotoGasm, Analog Diffusion a epiCRealism)

Stable Diffusion

(SD 1.5, SDXL)

- Free & open source
- Provozovatelná lokálně
- necenzurovaná
- Skrz finetuning umí +/- všechno-> komunitní, volně dostupné modely
- Obrovský ekosystém FOSS nástrojů
- Stojí na ní většina aplikací a služeb na AI obrázky (Leonardo.ai, Dreamstudio, Lexica…)

Čeho si všímat a jak poznat generované obrázky?



Alena Schillerová
(jako Vermeerova Dívka s perlou)
autor: David Laufer

Andrej Babiš
(jako Michelangelův David)
autorka: Melisa Çolakoglu

Zuzana Čaputová
(jako Ježíš)
autor: Dominik Jelínek

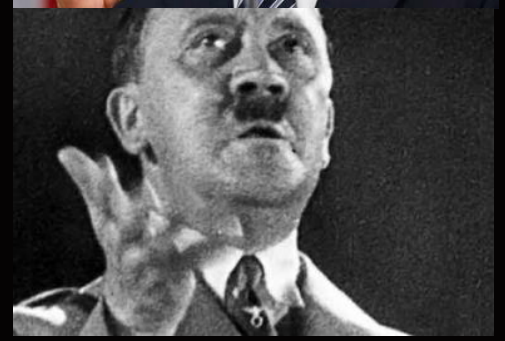
Klonování hlasu

Text-to-audio: Bark IIElevenLabs

Dataset hlasu -> tréning neuronovky -> výroba hlasu z textu

Audio-to-audio „voicechangery“

- Často už **v reálném čase**!
- Pro tréning modelu stačí <10 minut hlasu, někdy i podstatně méně.
- Audioknihu Karla Gotta si uděláte doma a možná s přesvědčivější intonací ☺



Nástroj: Retrieval-based-Voice-Conversion

Video

Autor: chaindrop (3/2023)
Nástroj: Modelscope text2video

Autor: Nicolas Neubert (11/2023)
Nástroj: Midjourney + Runwayml Gen2

Autor: Joshua Xu (8/2023)
Nástroj: HeyGen Instant Avatar

Autor: ColinandSamir (10/2023)
Nástroj: HeyGen Video Translate

Autor: shanef3d
Nástroj: EmberGen

„Petr Pavel konečně kápnul božskou“
Nástroj: Střih Videa® (6/2023)

Co z toho plyne?

- Kreativní profese čeká nejdivočejších 5 let v historii – i kdyby technologie zůstala na současné úrovni (focení, hudba, film, grafika…)
- Bez **důvěryhodného zdroje** nevěřte fotkám, nahrávkám, ani videu (to ale ve skutečnosti platilo i doteď).
- „Zaheslujte“ si blízké, protože vytvořit klon hlasu a používat ho jako voice changer při rozhovoru **v reálném čase** je jednoduché.
- Real-time deepfaky a filtry se **zatím** dají rozbít např. částečným zakrytím obličeje, pootočením… služby jako Instant Avatar zase neumožňují pohyb kamery / mluvčího.
- Sledujte vývoj AI a užijte si tu jízdu, žijeme v době, na kterou se bude vzpomínat.