Nova doba avtonomne umetne inteligence

Dr. Mladen Borovič, Fakulteta za elektrotehniko, računalništvo in informatiko, Univerza v Mariboru

REPUBLIKA SLOVENIJA
MINISTRSTVO ZA DIGITALNO PREOBRAZBO

REPUBLIKA SLOVENIJA
MINISTRSTVO ZA VISOKO ŠOLSTVO, ZNANOST IN INOVACIJE

REPUBLIKA SLOVENIJA
MINISTRSTVO ZA GOSPODARSTVO, TURIZEM IN ŠPORT

REPUBLIKA SLOVENIJA
MINISTRSTVO ZA OBRAMBO

ARIS
Javna agencija za znanstvenoraziskovalno in inovacijsko dejavnost Republike Slovenije

Od pogovornih sistemov do agentov

KCUI

Doba znanja

2022

Uporabni pogovorni sistemi

- ChatGPT
- Sledenje navodilom
- Generativna umetna inteligenca

2023

Konkurenčnost in sprejemanje

- Kopiloti (Microsoft, Google)
- Izboljšani modeli (GPT-4)
- Prvi agenti (AutoGPT)

Doba sklepanja

2024

Večmodalnost in večja kontekstna okna

- Obdelava slik in zvoka
- Ogromna kontekstna okna
- Vizualno “razumevanje”

2025

Logično načrtovanje, sklepanje in “razmišljanje”

- Obdelava po korakih
- Globoko raziskovanje
- Manj halucinacij

Doba agentov

2026

Avtonomni agenti

- Standardiziranje MCP
- Ogrodje OpenClaw
- Avtomatizacija procesov

UI v vlogi praktikanta

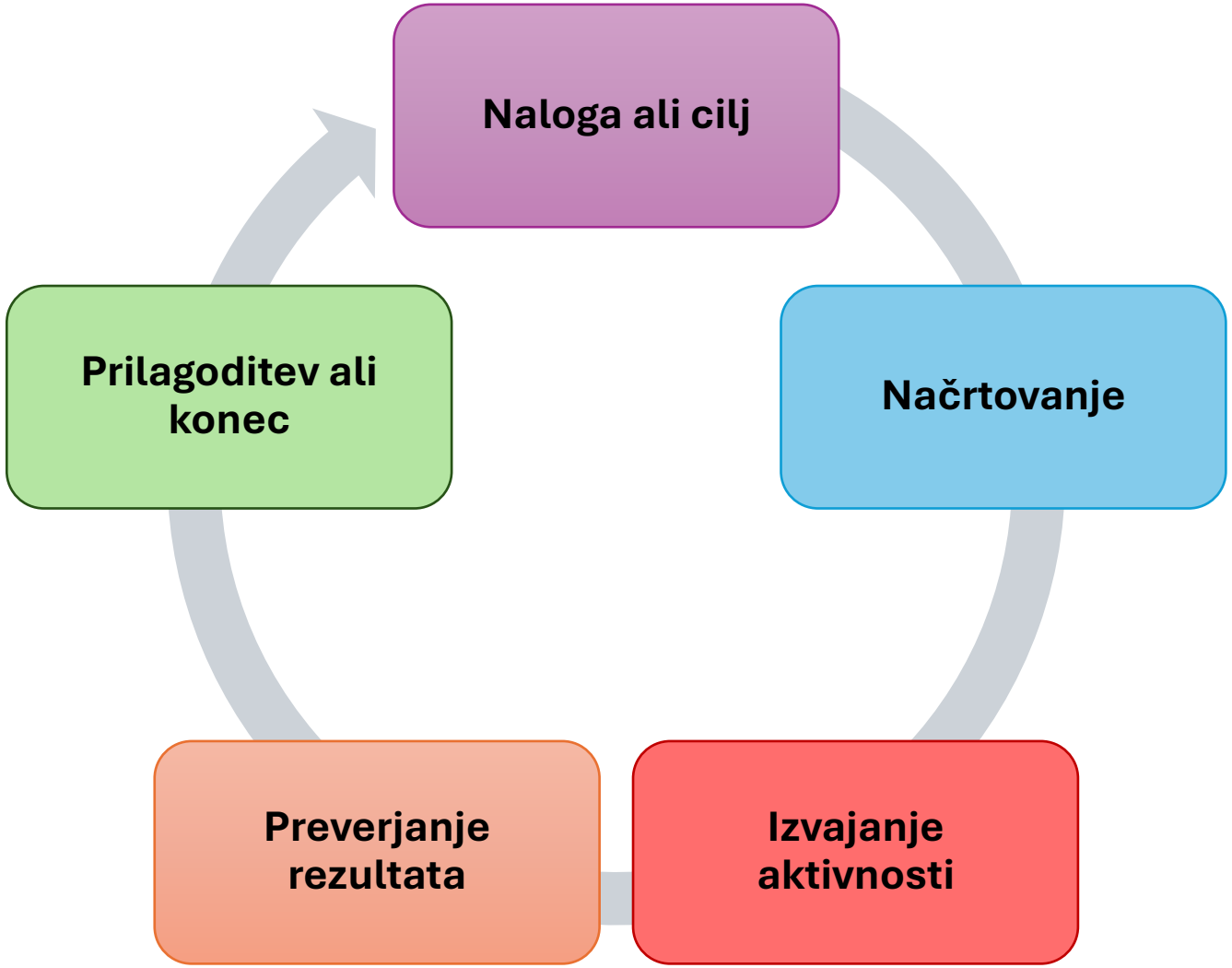
UI v vlogi svetovalca

UI v vlogi sodelavca

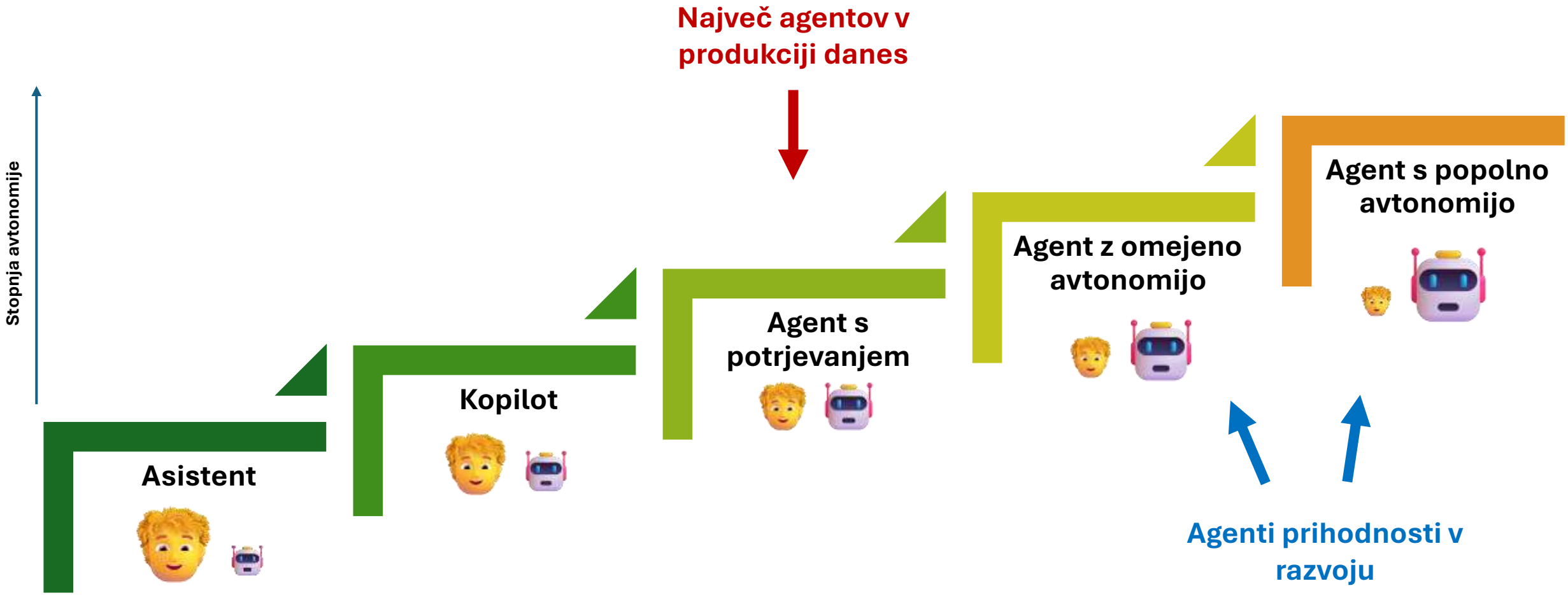
Agenti umetne inteligence

- Agente poganjajo modeli s sklepanjem (ang. reasoning models)
- Agenti imajo različne stopnje avtonomije
- Tehnologije za implementacijo

Claude
OpenAI
Gemini
deepseek
Qwen
MINIMAX
Mistral AI
Model Context Protocol
LangChain
n8n
CLAUDE CODE
zapier
OPENCLAW



Spekter avtonomije



Idealni procesi za avtomatizacijo

- Lastnosti idealnih procesov za agente UI
  - Jasno definirani koraki procesa
  - Dostopnost in povezljivost (npr. z MCP ali n8n)
  - Ponovljivost

- Primer: usklajevanje podatkov v računovodstvu
  - **Cilj** – uparjanje in organizacija nestrukturiranih ali delno strukturiranih podatkov
  - **Korak 1:** agent pridobi podatke iz CRM (uporaba n8n)
  - **Korak 2:** agent uporabi model s sklepanjem za uparjanje transakcij s pomanjkljivimi podatki (npr. Janez Novak z J. Novak)
  - **Korak 3:** agent označi podatke, ki jih ni uspešno uparil za kasnejšo obdelavo

Izbira pravega primera uporabe

- Kako se odločiti ali je proces smiselno avtomatizirati z agenti UI?
- Ali programska oprema, ki jo uporabljamo ponuja API za pridobivanje stanja in podatkov?
- Ali lahko proces opišemo z logičnimi koraki v obliki navodila ali specifikacij?
- Kakšne so posledice, če:
  - agent naredi napako (ali lahko ponastavimo stanje?)
  - človek ugotovi napako (ali lahko agenta ustavi/dopolni/popravi?)

Tveganja uporabe agentov

- Halucinacije in napake v rezultatu
  - Pri agentih to pomeni napačno akcijo/odločitev in posledično finančno izgubo
  - **Rešitev:** verifikacija v več korakih
    - npr. dodatni agenti za verifikacijo, nato preverja človek

- Kontekstno zanašanje (ang. context drift)
  - Procesi niso vedno statični; sčasoma se delovanje agenta lahko poslabša
    - Sprememba procesa, dela procesa, oblike vhoda (navodil) ali izhoda (rezultata)
    - Sprememba velikega jezikovnega modela v ozadju
  - **Rešitev:** redno vrednotenje in preverjanje ustreznosti rezultatov na realnih primerih

- Varnost
  - Nepričakovana in škodljiva skrita navodila (ang. malicious prompt injections)
  - Uhajanje občutljivih podatkov
  - **Rešitev:** izvajanje v kontroliranih okoljih (ang. sandboxing)
    - Uporaba tehnologij, ki podpirajo izvajanje v takšnih okoljih (npr. MCP in NemoClaw)

Odgovornost in etika

- Usklajevanje z EU AI Act in GDPR
- Transparentnost o uporabi agentov UI
- Človek v zanki (ang. human in the loop)
  - Človek **vedno** potrdi ključne odločitve (npr. zaposlovanje, odpuščanje, finančne transakcije)
- Pravna odgovornost
  - Trenutni trend v EU – podjetja so odgovorna za posledice agentovih akcij

Zaključek

- Agenti UI so tukaj in prinašajo pomembne spremembe
  - Manj nadziranja nalog in več upravljanja agentov ter njihovih delovnih tokov
  - Tehnologije kot so MCP, n8n, OpenClaw in zmogljivi novi modeli omogočajo uporabo v praksi

- Tri strategije za začetek uporabe agentov UI
  - Manjši procesi z nizkimi tveganji
  - Vzpostavitev internih varovalk in postopkov za verifikacijo delovanja agentov
  - Postopno povečevanje stopnje avtonomije agentov

- Cilj uporabe agentov UI ni nadomestiti ljudi, temveč avtomatizirati ponavljajoče se delo in uporabiti znanje obstoječih zaposlenih za bolj kompleksne naloge.

Hvala za pozornost

- **Kontakt:**
  - mladen.borovic@um.si
  - ☎ +386 2 220 7460
  - LinkedIn

- **Raziskovalna in interesna področja:**
  - Aplikacije umetne inteligence
  - Veliki jezikovni modeli
  - Veliki večmodalni modeli
  - Priporočilni sistemi in iskalniki
  - Obdelava naravnega jezika
  - Detekcija podobnih vsebin
  - Visokozmogljivo računalništvo (HPC)

dr. Mladen Borovič
Univerza v Mariboru
Fakulteta za elektrotehniko, računalništvo in informatiko
Inštitut za računalništvo
Laboratorij za heterogene računalniške sisteme

REPUBLIKA SLOVENIJA
MINISTRSTVO ZA DIGITALNO PREOBRAZBO

REPUBLIKA SLOVENIJA
MINISTRSTVO ZA VISOKO ŠOLSTVO, ZNANOST IN INOVACIJE

REPUBLIKA SLOVENIJA
MINISTRSTVO ZA GOSPODARSTVO, TURIZEM IN ŠPORT

REPUBLIKA SLOVENIJA
MINISTRSTVO ZA OBRAMBO

ARIS Javna agencija za znanstvenoraziskovalno in inovacijsko dejavnost Republike Slovenije

Ko umetna inteligenca postane vaša sodelavka – ne vaš projekt

Vanja Belec, Pošta Slovenije d.d.

REPUBLIKA SLOVENIJA
MINISTRSTVO ZA DIGITALNO PREOBRAZBO

REPUBLIKA SLOVENIJA
MINISTRSTVO ZA VISOKO ŠOLSTVO, ZNANOST IN INOVACIJE

REPUBLIKA SLOVENIJA
MINISTRSTVO ZA GOSPODARSTVO, TURIZEM IN ŠPORT

REPUBLIKA SLOVENIJA
MINISTRSTVO ZA OBRAMBO

ARIS
Javna agencija za znanstvenoraziskovalno in inovacijsko dejavnost Republike Slovenije