

# Lekcije preteklosti za dobo umetne inteligence

**Kaj nas lahko na algoritmihi zasnovane  
informacijske rešitve naučijo za dobo  
umetne inteligence**

mag. Maja Hmelak

Gradivo je last Slovenskega inštituta za revizijo in je predmet avtorske zaščite in drugih oblik zaščite intelektualne lastnine. Prepovedano je kakršnokoli reproduciranje, razen izključno za osebno uporabo in v nekomercialne namene, pri čemer se morajo ohraniti vsa opozorila o avtorskih ali drugih pravicah, zato se ne smejo prepisovati, razmnoževati ali kako drugače razširjati. Naveden mora biti tudi vir.

## Kako je vrhovno sodišče pristopilo k informatizaciji sodišč?

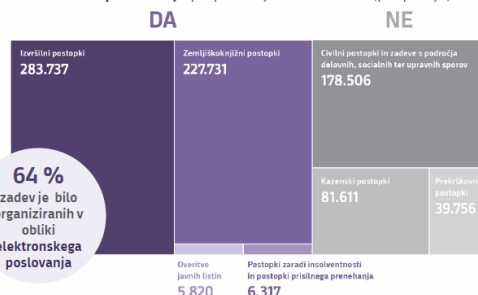
### DOLGOROČNO NAČRTOVANJE informacijske podpore

- Vrhovno sodišče je sledilo **dolgoročni strategiji** in **postopoma** podprlo ključne vidike delovanja sodišč.
- Strategija se je osredotočala na **čim večjo podprtost obsežnih, ponovljivih nalog** na primer zemljiškoknjižne in izvršilne postopke. Te naloge je vrhovno sodišče organiziralo tudi kot **elektronsko poslovanje**.
- Vrhovno sodišče je informacijsko podprlo tudi številne vidike **upravnega poslovanja sodišč**, na primer poslovno poročanje in centralizacijo tiskanja in odpreme pošte.
- Nekatera področja, med drugim civilni postopki, zadeve s področja delovnih, socialnih in upravnih sporov, kazenski postopki in prekrškovni postopki, še vedno temeljijo na fizični dokumentaciji.

### ELEKTRONSKO POSLOVANJE sodišč

- Notarji, odvetniki, stečajni upravitelji in drugi, lahko **sodne zadeve** pričnejo in vodijo preko **spleta**.
- Sodišča obravnavajo zadeve ob visoki stopnji **avtomatizacije**.
- Sodišča uporabljajo **centralizirane poštnne storitve**, vključno z elektronskim vročanjem.

#### Elektronsko poslovanje po področjih dela sodišč (povprečje)



### razvoju informacijske podpore

upoštevalo načela **enotnega razvojnega okolja** pripravih rešitev in standardnih formatov.

litve sodišč so praviloma **povezljive**, pomembne uporabljajo v **več rešitvah**, sodišča pa za delo z dokumenti plačilo pisarniškega orodja OpenOffice.

Računsko sodišče Republike Slovenije – revizijsko poročilo Zagotavljanje informacijske podpore delovanju sodišč

REPUBLIKA SLOVENIJA  
RAČUNSKO SODIŠČE

REVIZIJSKO POROČILO  
Zagotavljanje informacijske podpore delovanju sodišč



## Težave zaposlenih s predpripravljenimi odločbami v informacijski podpori CSD



**Zaradi tehničnih napak** so morali zaposleni na CSD **večkrat** izvajati **iste naloge**, pri nekaterih nalogah pa so morali izvajati **dodatne korake** in celo **ročno popravljati** izdane odločbe.

slovnicične napake

niso enotno uporabljale izrazov

včasih kontradiktorne

pogosto brez ustrezne obrazložitve

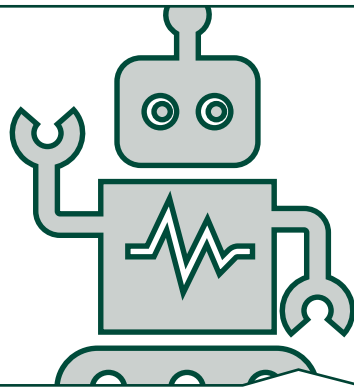
pogosto navajale neveljavne ali nerelevantne pravne podlage

včasih večkrat ponavljale iste odstavke

pogosto niso vsebovale osnovnih informacij o izračunu ali trajanju pravice



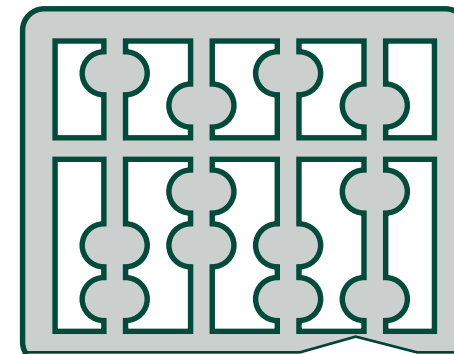
# Kaj ta teden imenujemo umetna inteligenca



Strojni vid



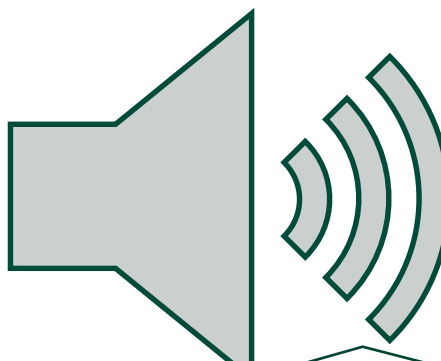
Generativna umetna  
inteligenca



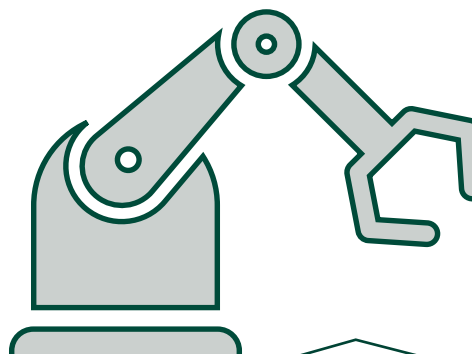
Napovedna analitika



Priporočilni sistemi



Obdelava naravnega  
jezika

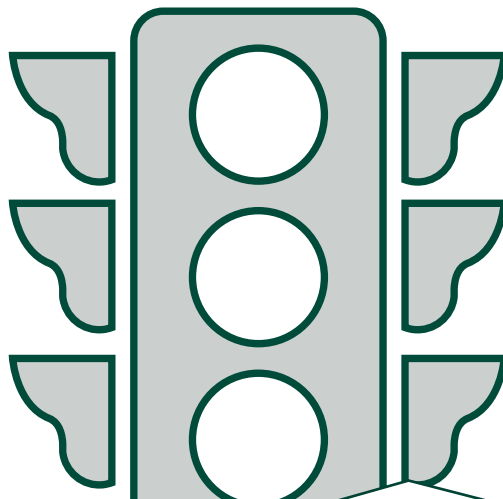


Avtonomni sistemi  
odločanja

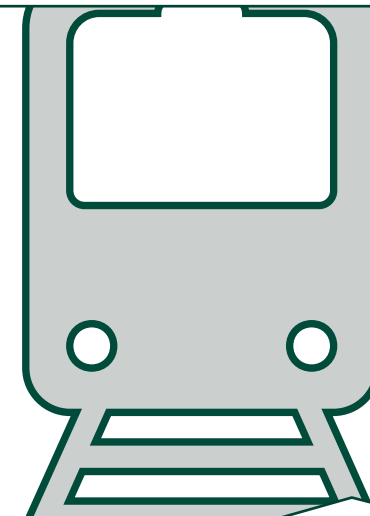
# Kje vse že srečamo umetno inteligenco



Zaznavanje prometnih prekrškov –  
Pune Indija in Avstralija



Samodejno posredovanje prijav  
napak na javni infrastrukturi - Tajska



Spremljanje mikropoškodb prog z  
merjenjem tresljajev na nameščenih  
mobilnih telefonih - ZDA NY

NIST

*Okvir za upravljanje tveganj umetne inteligence*

NIST

*Okvir za upravljanje tveganj umetne inteligence:  
Profil generativne umetne inteligence*

Evropska  
komisija

*Etična navodila za zaupanja vredno umetno  
inteligenco*

# Etična navodila za zaupanja vredno umetno inteligenco:

## 4 ključna etična načela

**Spoštovanje  
človekove  
avtonomije**

Ali informacijska rešitev zagotavlja spoštovanje pravic posameznika?

**Preprečevanje  
škode**

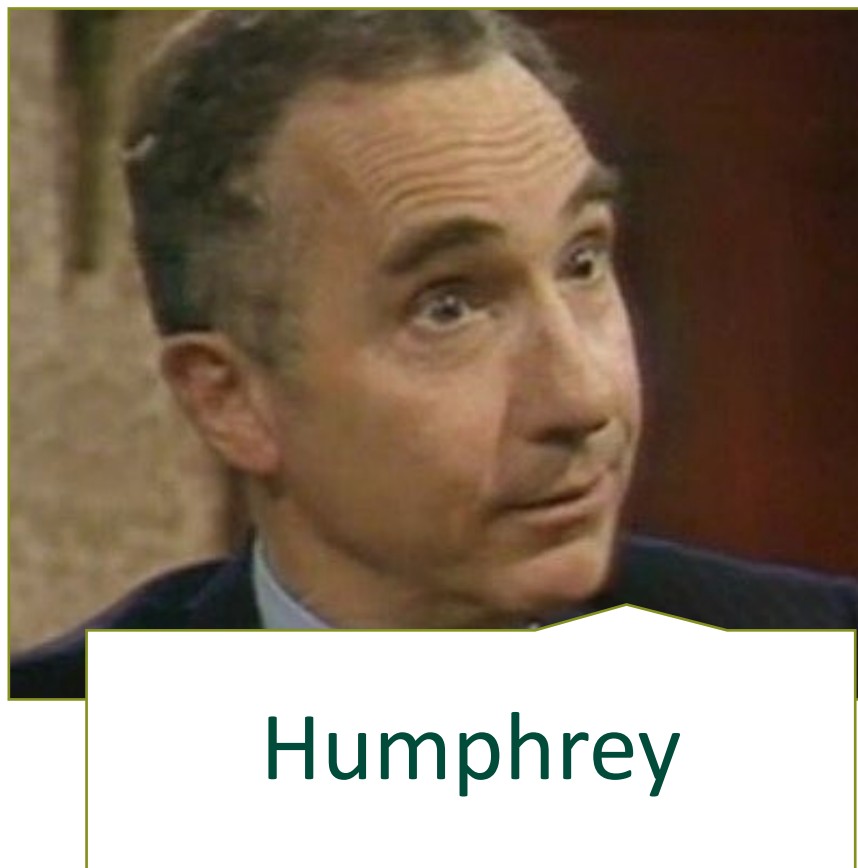
Ali je zagotovljeno, da informacijska rešitev ne povzroča škode?

**Pravičnost**

Ali informacijska rešitev zagotavlja pravično obravnavo?

**Transparentnost**

Ali je mogoče v celoti pojasniti, kako informacijska rešitev deluje?



## Ključni vidiki 4 etičnih načel

### **Spoštovanje človekove avtonomije:**

Ali javni uslužbenci ohranijo možnost samostojne presoje in odločanja, ko uporabljajo orodja, kot sta Copilot in Humphrey, ali pa se njihovo delo preveč zanaša na avtomatizirane predloge umetne inteligence?

### **Preprečevanje škode:**

Ali obstajajo postopki, ki preprečujejo, da bi umetna inteligenca ustvarjala napačne ali pristranske informacije v uradnih dokumentih, analizah ali zapisnikih?

### **Pravičnost:**

Ali orodje za zbiranje povratnih informacij obravnava vse prispevke javnosti enakovredno – ne glede na ton, dolžino, način izražanja ali jezik – ter omogoča, da se mnenja državljanov vključijo na pravičen in pregleden način?

### **Razložljivost:**

Ali je mogoče pojasniti, na kakšen način umetna inteligenca povzema vsebine ali analizira odzive, in kdo nosi odgovornost za končne interpretacije in poročila, ki jih sistem ustvari?





## Ključni vidiki 4 etičnih načel

### **Spoštovanje človekove avtonomije:**

Ali ima posameznik možnost izvedeti, da je bil označen, in razumeti, zakaj?

### **Preprečevanje škode:**

So bili vzpostavljeni postopki za preprečevanje napačnih označitev in neupravičenih ukrepov?

### **Pravičnost:**

Ali sistem obravnava vse skupine prebivalstva enako, ne glede na socialni ali etnični položaj?

### **Razložljivost:**

Je mogoče pojasniti, kako sistem sprejema odločitve, in kdo je zanje odgovoren?

$$P_{kj} = (1 - r_j) * C_{kj} + r_j * (C_{kj} + q_{kj} - p_{kj})$$

- $P_{kj}$  – predviden delež dijakov na šoli  $j$ , ki bodo dosegli oceno  $k$
- $C_{kj}$  – zgodovinski delež dijakov na šoli  $j$ , ki so v zadnjih treh letih dosegli oceno  $k$
- $q_{kj}$  – predviden delež trenutnih dijakov na šoli  $j$ , za katere se pričakuje, da bodo dosegli oceno  $k$ , glede na njihove predhodne dosežke
- $p_{kj}$  – predviden delež prejšnjih generacij dijakov na šoli  $j$ , za katere se je pričakovalo, da bodo dosegli oceno  $k$ , glede na njihove predhodne dosežke
- $r_j$  – delež dijakov na šoli  $j$ , za katere so na voljo podatki o prejšnjih dosežkih

## Explainer

### Ofqual's A-level algorithm: why did it fail to make the grade?

There is a lot we can learn from the algebraic symbols used to determine results in England

● [A university vice-chancellor's diary of A-level chaos](#)



## Ključni vidiki 4 etičnih načel

### **Spoštovanje človekove avtonomije:**

Ali je bil vsak dijak obravnavan kot posameznik z lastnimi dosežki in napredkom, ali pa je algoritem ocene dodeljeval pavšalno – na podlagi povprečij šole in preteklih rezultatov drugih?

### **Preprečevanje škode:**

Ali so bili pred uvedbo algoritma ocenjeni možni negativni učinki napačno dodeljenih ocen – na primer izguba možnosti za nadaljnje izobraževanje – in ali so obstajali postopki za njihovo odpravo?

### **Pravičnost:**

Ali je sistem zagotavljal enako obravnavo vseh dijakov, ne glede na to, ali so obiskovali šole z nižjo zgodovinsko uspešnostjo, in ali je preprečeval posredno diskriminacijo socialno šibkejših skupin?

### **Razložljivost:**

Ali je bilo mogoče posameznemu dijaku ali učitelju razumljivo pojasniti, kako je bila ocena določena, in ali je bila omogočena pritožba zoper nepravično oceno?



Shutterstock/Oria

Copy link  
Email  
X (Twitter)  
Bluesky  
Facebook  
LinkedIn  
WhatsApp  
Print

Justice systems around the world are using artificial intelligence (AI) to assess people with criminal convictions. These AI technologies rely on machine learning algorithms and their key purpose is to predict the risk of reoffending. They influence decisions made by the courts and prisons and by parole and probation officers.

This kind of tech has been an intrinsic part of the UK justice system since 2001. That was the year a risk assessment tool, known as Oasys (Offender Assessment System), was introduced and began taking over certain tasks from probation officers.

Yet in over two decades, scientists outside the government have not been permitted access to the data behind Oasys to independently analyse its workings and assess its accuracy – for example, whether the decisions it influences lead to fewer offences or reconvictions.

## Inquest to examine probation's handling of Killamarsh murderer Damien Bendall

CALENDAR | CHESTERFIELD | CRIME AND COURTS | SHEFFIELD | Tuesday 18 April 2023 at 10:12pm



An inquest will investigate how a violent man was deemed to pose a low risk of serious harm by probation services before going on to murder his pregnant partner and three children.

Damien Bendall had a 17-year history of serious and violent crime before killing 35-year-old Terri Harris and her two children John Paul Bennett, 13, and Lacey Bennett, 11 and Lacey's friend Connie Gent who was on a sleepover at their home in Killamarsh near [Sheffield](#).

## Ključni vidiki 4 etičnih načel

### **Spoštovanje človekove avtonomije:**

Ali ima strokovno osebje pri uporabi sistema OASys dovolj možnosti, da uveljavi lastno strokovno presojo, ali pa se pri odločanju o tveganju ponovitve kaznivih dejanj preveč zanaša na rezultat, ki ga izračuna informacijska rešitev?

### **Preprečevanje škode:**

Ali so v sistemu vzpostavljeni postopki, ki zagotavljajo pravočasno posodabljanje podatkov o obsojencih in s tem preprečujejo napačne ocene tveganja, ki bi lahko vodile v resne posledice za varnost ljudi?

### **Pravičnost:**

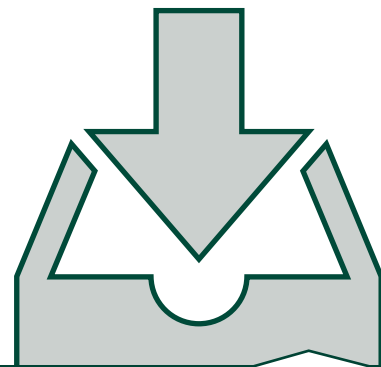
Ali sistem enako ocenjuje vse obsojence, ne glede na njihovo preteklost, socialni status ali vedenjske značilnosti, ter ali omogoča dovolj individualne obravnave pri odločanju o pogojni izpustitvi?

### **Razložljivost:**

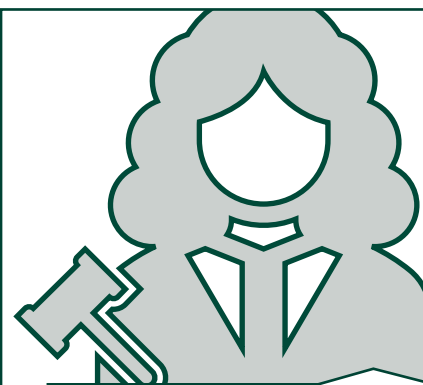
Ali je mogoče jasno pojasniti, kako sistem kombinira različne dejavnike tveganja v končno oceno, in ali so ti postopki dovolj pregledni, da omogočajo zunanji nadzor in odgovornost?



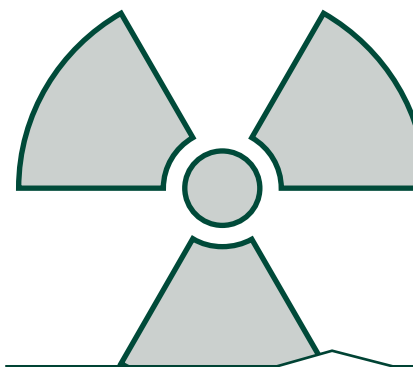
Tehnologija še ne zagotavlja  
pravičnosti



Vhodni podatki odločajo o  
izhodih



Algoritmi ne morejo  
nadomestiti človeške presoje



Uporaba tehnologij UI ne  
ustvarja novih tveganj, lahko  
pa poveča obstoječa

# Hvala za pozornost!

