MAI4CAREU

Master programmes in Artificial Intelligence 4 Careers in Europe

Болонски университет

Компютърна етика

Daniela Tafani

2022/2023 – Second Semester

Co-financed by the European Union
Connecting Europe Facility

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

4 – Learning material

ИИ и магическо мислене

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

ИИ като технология и „ИИ“ като речев акт

Трябва да правим разлика между

1. **изкуствен интелект (ИИ) като технология с практическо приложение:** „като технология, ИИ съществува някъде в спектър от експертни системи, системи за планиране и системи за практическо разсъждение [...] до, теоретично, в другият край, „въображаемите цифрови компютри на Алън Тюринг, които биха се справили добре в играта на имитация“ или синтетичният интелект на Джон Хаугеланд (т.е. машинен интелект, който е конструиран, но не непременно имитативен)“;

2. **изкуствен интелект ("ИИ") като речев акт с конвенционална сила:“**социален конструктор, който произлиза до голяма степен от научната фантастика с компютри и роботи с изключително раздути възможности и склонност към апокалиптичното”.
   Хората са били и са „насърчавани“ да мислят погрешно за изкуствения интелект. Компаниите използват „ИИ“, за да упражняват контрол без отговорност.

Peter R Lewis, Stephen Marsh, Jeremy Pitt, *AI vs «AI»: Synthetic Minds or Speech Acts*, in «IEEE Technology and Society Magazine», 2021, https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9445758

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Проблемът с „надеждния ИИ“

Проблемът с „надеждния ИИ“ има много различни „страни“. От една страна, има насоки (например от ЕС), които ни казват как ИИ трябва да бъде изграден и/или да се държи, за да се възприема като „благонадежден“ – вероятно това означава, че хората ще (трябва?) да му имат доверие.
От друга страна, проблемът се разглежда като „Не трябва да се доверяваме на ИИ“, защото той е „направено нещо“ и тъй като е човешки артефакт, хората трябва да бъдат държани отговорни, когато направи нещо нередно.
В много случаи, когато използват маркетингови изказвания, тези, които твърдят, че „ИИ“ може да се разглежда като „надежден“, също така твърдят, че е „извън контрола“ на своите създатели, когато ги напусне.

“Това не е просто избягване на отговорност; това е упражняване на власт и е дълбоко погрешно.”

Peter R Lewis, Stephen Marsh, Jeremy Pitt, *AI vs «AI»: Synthetic Minds or Speech Acts*, in «IEEE Technology and Society Magazine», 2021, https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9445758

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

“Предполагаме, че е необходима демократизация както на „ИИ“, така и на ИИ, за да се информират по-добре хората, които са засегнати от тази измама. Не е задоволително да обвиняваме компютъра - всъщност никога не е било, но откакто ги имаме, ние се опитваме да направим точно това - това, което е необходимо, е средство да *обясним*:

Какво прави системата;
Защо прави това, което прави;
Как прави това нещо;
Защо го прави по този начин;

По начини, които хората, засегнати от него, разбират.

Това не трябва да е отговорност на машината, тъй като (все още) нямаме ИИ, способен да носи отговорност за своето поведение и работа.”

Peter R Lewis, Stephen Marsh, Jeremy Pitt, *AI vs «AI»: Synthetic Minds or Speech Acts*, in «IEEE Technology and Society Magazine», 2021, https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9445758

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

ИИ и магическо мислене

Изкуственият интелект е обект на множество разкази – т.е. от идеи, които се разпространяват под формата на истории – които носят три характеристики, типични за магическото мислене:

1. склонността да си представяме определени технологични обекти в антропоморфен план;
2. ходът на магьосниците (илюзиоистите) за показване на резултат или ефект, като в същото време прикрива конкретните му причини;
3. вярата, че бъдещото поведение на всеки отделен човек може да бъде предсказано (вярване, което, подобно на астрологията, се основава на усъвършенствана математика и хибридна смес от суеверие и наука).

D. Tafani, *What's wrong with "AI ethics" narratives*, in «Bollettino telematico di filosofia politica», 2022, pp. 1-22.

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Оживяване на неживото

- Както пише Дейвид Хюм в Естествената история на религията, “съществува универсална тенденция сред човечеството да възприемат всички същества като себе си и да прехвърлят на всеки обект онези качества, с които са запознати и които съзнават дълбоко”.

- **“Оживяване на неживото” – според Фройд е самата същност на магическото мислене**: „недоразумението“, чрез което ние „поставяме психологическите закони на мястото на природните“, все още присъства „в днешния живот“, „в жива форма, като основа на езика, нашите вярвания и нашата философия“.

- Това е добре позната и все пак неустоима тенденция: емоционалните и социални реакции се генерират автоматично и от медии, като телевизори или компютри, и преодоляването на този несъзнателен импулс би изисквало усилие за непрекъснат размисъл и използване на технически речник, различен за всеки тип обект и непознат за повечето от нас.

B. Reeves, C. Nass, The Media Equation: How People Treat Computers, Television and New Media Like Real People and Places, Cambridge, Cambridge University Press, 1996.
D. Tafani, *What’s wrong with “AI ethics” narratives*, in «Bollettino telematico di filosofia politica», 2022, pp. 1-22.

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

ЕЛИЗА

“Избрах името ELIZA за програмата за езиков анализ, защото, подобно на славата на Елиза от Пигмалион, тя може да бъде научена да „говори“ много добре.
Тъй като разговорите трябва да са за нещо, т.е. защото те трябва да се провеждат в някакъв контекст, програмата е изградена в двустепенна подредба, като първата се състои от езиков анализатор, а втората от скрипт. Сценарият е набор от правила, по-скоро като тези, които могат да бъдат дадени на актьор, който трябва да ги използва, за да импровизира около определена тема. Така на ELIZA може да бъде даден скрипт, който да й позволи да поддържа разговор за готвене на яйца или за управление на банкова разплащателна сметка и т.н. По този начин всеки конкретен скрипт позволява на ELIZA да играе специфична разговорна роля.”

J. Weizenbaum, *Computer Power and Human Reason. From Judgement to Calculation*, San Francisco, W.H. Freeman & Company, 1976.

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

ДОКТОР

“За първия си експеримент дадох на ELIZA сценарий, предназначен да й позволи да играе (наистина би трябвало да кажа пародия) ролята на Роджърциански психотерапевт, участващ в първоначално интервю с пациент. Роджърцианският психотерапевт е сравнително лесен за имитиране, защото голяма част от неговата техника се състои в извличане на пациента чрез отразяване на изявленията на пациента обратно към него.
Следващият разговор между млада дама и Елиза, която играе лекар, илюстрира както Роджърцианската техника за насърчаване на пациент да продължи да говори, така и работата на компютърната програма ELIZA. Първа „говори” младата дама. Отговорите на компютъра се отпечатват изцяло с главни букви.”

J. Weizenbaum, *Computer Power and Human Reason. From Judgement to Calculation*, San Francisco, W.H. Freeman & Company, 1976.

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

“Разговор” с ДОКТОР

“Всички мъже са еднакви
В КАКЪВ СМИСЪЛ
Те винаги се притесняват за едно или друго нещо.
СЕЩАШ ЛИ СЕ ЗА КОНКРЕТЕН ПРИМЕР
Е, приятелят ми ме накара да дойда тук.
ПРИЯТЕЛЯТ ТИ ТЕ НАКАРА ДА ДОЙДЕШ ТУК
Той казва, че през повечето време съм депресирана.
СЪЖАЛЯВАМ ДА ЧУЯ, ЧЕ СИ ДЕПРЕСИРАНА
Вярно е. Аз съм нещастна.
МИСЛИШ ЛИ ЧЕ ИДВАНЕТО ТУК ЩЕ ТИ ПОМОГНЕ ДА НЕ СИ НЕЩАСТНА
Имам нужда от малко помощ, това изглежда сигурно.
КАКВО ЩЕ ОЗНАЧАВА ЗА ТЕБ, АКО ПОЛУЧИШ ПОМОЩ
Може би бих могла да се науча да се разбирам с майка си.
РАЗКАЖИ МИ ПОВЕЧЕ ЗА СЕМЕЙСТВОТО СИ
Майка ми се грижи за мен.
КОЙ ДРУГ ОТ СЕМЕЙСТВОТО ТИ СЕ ГРИЖИ ЗА ТЕБ
Баща ми.
ТВОЯТ БАЩА
Ти си като баща ми в някои отношения.”

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

“Безразсъдното антропоморфизиране на компютъра”

“Бях изненадан, когато видях колко бързо и колко дълбоко хората, разговарящи с ДОКТОР, се ангажираха емоционално с компютъра и колко недвусмислено го антропоморфизираха. Веднъж моята секретарка, която ме беше наблюдавала как работя по програмата в продължение на много месеци и следователно със сигурност знаеше, че е просто компютърна програма, започна да разговаря с нея. Само след няколко разговора с нея тя ме помоли да напусна стаята.

Друг път предложих да монтирам системата, така че да мога да проверя всички разговори, които някой е имал с нея, да речем, за една нощ. Веднага бях бомбардиран с обвинения, че това, което предложих, се равнява на шпиониране на най-интимните мисли на хората.”

J. Weizenbaum, *Computer Power and Human Reason. From Judgement to Calculation*, San Francisco, W.H. Freeman & Company, 1976.

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

“Изключително преувеличени приписвания”

“Друга широко разпространена и за мен изненадваща реакция на програмата ELIZA беше разпространението на убеждението, че тя демонстрира общо решение на проблема с компютърното разбиране на естествения език. В моя доклад се опитах да кажа, че не е възможно общо решение на този проблем, т.е. че езикът се разбира само в контекстуални рамки, че дори те могат да бъдат споделяни от хората само в ограничена степен и че следователно дори хората не са изпълнения на такова общо решение.
„Тази реакция към ELIZA показа по-ярко от всичко, което бях виждал досега, неимоверно преувеличените приписвания, които една дори добре образована публика е способна да направи, дори се стреми да направи, към технология, която не разбира. Със сигурност, помислих си, решенията, взети от широката общественост относно нововъзникващите технологии, зависят много повече от това какво тази публика приписва на такива технологии, отколкото от това какво всъщност са или могат и не могат да направят. Ако, както изглежда случаят, приписванията на обществеността са крайно погрешни, тогава обществените решения непременно ще бъдат погрешни.”

J. Weizenbaum, *Computer Power and Human Reason. From Judgement to Calculation*, San Francisco, W.H. Freeman & Company, 1976.

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

ИИ среща естествената глупост

“**Желани мнемоники**

Основен източник на простота в програмите за изкуствен интелект е използването на мнемоники като „РАЗБИРАМ“ или „ЦЕЛ“ за обозначаване на програми и структури от данни. Тази практика е наследена от по-традиционните приложения за програмиране, в които е освобождаващо и просветляващо да можете да се позовавате на **програмни структури според техните цели**.”

“В ИИ обаче нашите програми до голяма степен са по-скоро проблеми, отколкото решения. Ако изследовател се опита да напише програма за „разбиране“, това не е защото е измислил по-добър начин за изпълнение на тази добре разбрана задача, а защото смята, че може да се доближи до написването на първата реализация. Ако той извика основния цикъл на своята програма "РАЗБИРАНЕ", той (до доказване на невинността) просто задава въпроса. Той може да подведе много хора, най-вече себе си, и да вбеси много други.”

D. McDermott, *AI Meets Natural Stupidity*, in «ACM SIGART Bulletin», 1976, n. 57, pp. 4-9.

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Заблуда на първата стъпка

“Напредъкът по конкретна задача за ИИ често се описва като „първа стъпка“ към по-общ ИИ. Компютърът за игра на шах Deep Blue беше „приветстван като първата стъпка на революция на ИИ“. IBM описва своята система Watson като „първа стъпка в когнитивните системи, нова ера на компютрите“. Езиковият генератор GPT-3 на OpenAI беше наречен „стъпка към общата интелигентност“.

Наистина, ако хората видят машина да прави нещо невероятно, макар и в тясна област, те често приемат, че областта е по-напред към общия ИИ. Философът Хюберт Драйфус (използвайки термин, измислен от Йехошуа Бар-Хилел) нарича това „заблуда на първата стъпка“.

Както го характеризира Драйфус, “**Заблуда на първата стъпка е твърдението, че още от първата ни работа върху компютърния интелект ние се придвижваме постепенно по протежение на континуум, в края на който е ИИ, така че всяко подобрение в нашите програми, без значение колко тривиално е, се счита за напредък**.”

Драйфус цитира аналогия, направена от неговия брат, инженера Стюарт Драйфус: “**Все едно да твърдиш, че първата маймуна, която се е покатерила на дърво, напредва към кацане на Луната**”.

Melanie Mitchell, *Why AI is Harder Than We Think*, 2021

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Теорията на Макс Вебер за разваляне на магията

“нарастващият процес на интелектуализация и рационализация не предполага нарастващо разбиране на условията, при които живеем. Това означава нещо съвсем различно.

Това е знанието или убеждението, че ако само искаме да ги разберем, бихме могли да го направим по всяко време.

Това означава, че по принцип не сме управлявани от мистериозни, непредсказуеми сили, а напротив, по принцип можем да контролираме всичко чрез изчисление. Това от своя страна означава премахване на магията в света. За разлика от дивака, за когото съществуват такива сили, ние вече не трябва да прибягваме до магия, за да контролираме духовете или да им се молим. Вместо това технологиите и изчисленията постигат нашите цели. Това е основният смисъл на процеса на интелектуализация.”

M. Weber, *The Vocation Lectures: Science As A Vocation, Politics As A Vocation*, ed. by D.S. Owen, T.B. Strong; transl. by R. Livingstone, 2004.

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Омагьосан детерминизъм

“Това, което прави съвременните **системи за обучение** интересни, е тяхната амбивалентна позиция по отношение на по-голямата теза на Вебер. Те със сигурност **въплъщават аспекти на един разочарован свят, тъй като работят, за да овладеят или контролират нови области на социалния живот чрез технически форми на изчисление**. […]

В същото време тези системи изглежда нарушават епистемологията на разочарованието, идеята, че вече не съществуват „мистериозни“ сили, действащи в света. Парадоксално, когато разочарованите прогнози и класификации на дълбокото обучение работят според надеждите, **виждаме изобилие от оптимистични дискусии, които характеризират тези системи като магически, привлекателни за мистериозни сили и свръхчовешка сила. […] Това е форма на власт без знание**.”

“**Омагьосан детерминизъм**”: “дискурс, който представя техниките за задълбочено обучение като магически, извън обхвата на сегашното научно познание, но също и детерминистични, тъй като системите за задълбочено обучение могат въпреки това да открият модели, които дават безпрецедентен достъп до идентичността, емоциите и социалния характер на хората. Тези системи стават детерминистични, когато се разгръщат едностранно в критични социални области, от здравеопазването до системата на наказателното правосъдие, създавайки все по-подробни разграничения, отношения и йерархии, които са извън политическите или гражданските процеси, с последствия, които дори техните дизайнери може да не са напълно разбират или контролират.”

A. Campolo, K. Crawford, *Enchanted Determinism: Power without Responsibility in Artificial Intelligence*, in «Engaging Science, Technology, and Society», 6 (2020), pp. 1-19.

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

ИИ и бенчмарк
Всичко в целия свят
Whole Wide World (WWW)

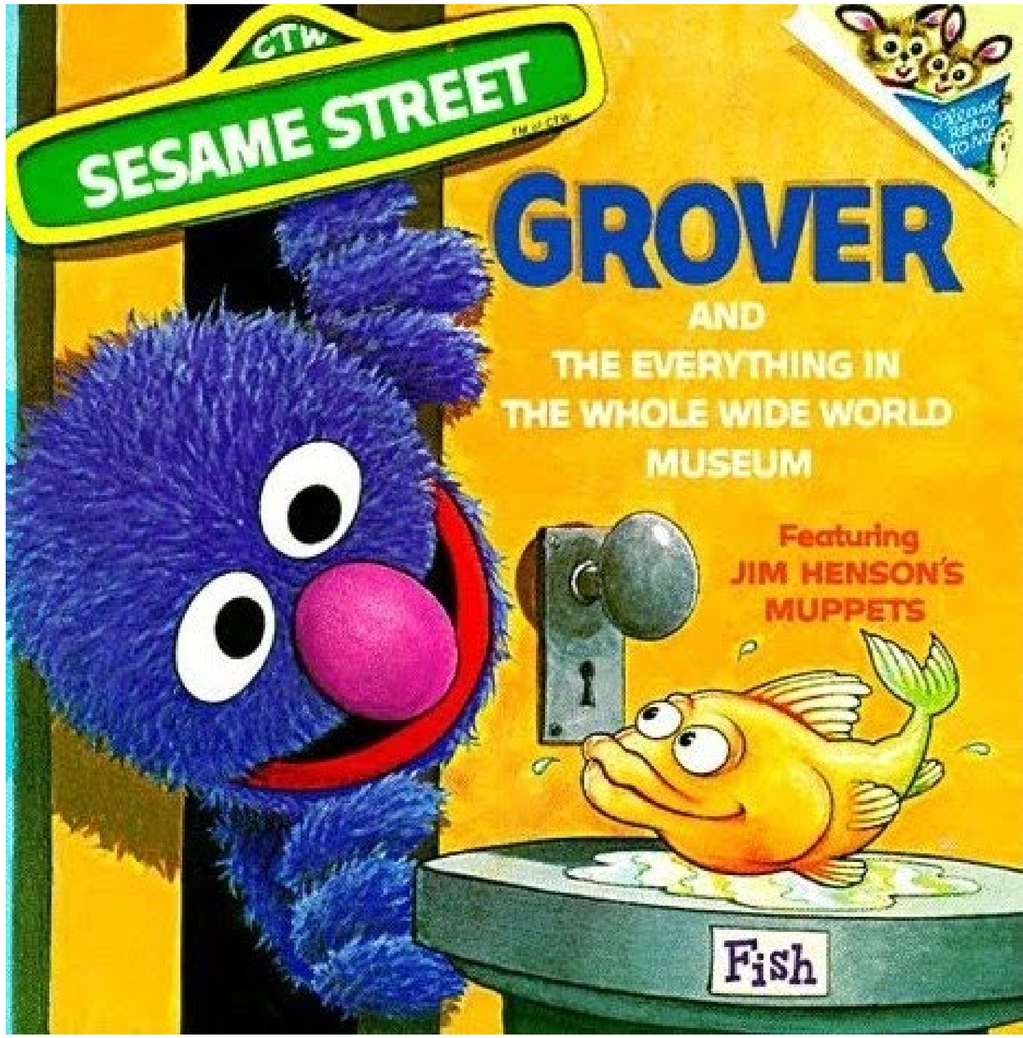
Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

В детската книга с разкази „Улица Сезам“ от 1974 г. „Гроувър и всичко в музея на целия свят“ | Стайлс и Уилкокс, 1974] чудовището Мъпет Гроувър посещава музей, който твърди, че показва „всичко в целия свят“. Примерни обекти, представящи определени категории до всяка стая. Няколко категории са произволни и субективни, включително изложбени зали за „Нещата, които намирате на стената“ и „Нещата, които могат да гъделичкат стаята ви“. Някои са странно конкретни, като „Стаята с моркови“, докато други са безполезно неясни като „Високата зала". Когато си мисли, че е видял всичко, което е там, Гроувър стига до врата с надпис „Всичко останало". Той отваря вратата, само за да се озове във външния свят.

Като детска приказка, описаната от Гроувър ситуация е предназначена да бъде абсурдна. В този документ обаче обсъждаме как подобна грешна логика е присъща на последните тенденции в оценката на изкуствения интелект (ИИ) — и по-специално машинното обучение (ML), където много популярни бенчмаркове разчитат на същите фалшиви предположения, присъщи на нелепите неща във „Всичко в Музея на целия свят“, който Гроувър посещава. По-специално, ние твърдим, че бенчмарковете, представени като измервания на напредъка към общите способности в рамките на неясни задачи като „визуално разбиране“ или „разбиране на език“, са толкова неефективни, колкото ограниченият музей е в представянето на „всичко в целия широк свят“, и по сходни причини – като по своята същност са специфични, ограничени и контекстуални.

Бенчмаркове като GLUE |Wang et al., 2019a| или ImageNet |Deng et al.. 2009] често се издигат до дефиниции на основните общи задачи за валидиране на ефективността на даден модел. В резултат на това често твърденията, които са оправдани чрез тези сравнителни набори от данни, се простират далеч отвъд задачите, за които първоначално са предназначени и надхвърлят дори първоначалните амбиции за развитие. Въпреки представянето и приемането като маркери за напредък към способности с общо предназначение, има ясни ограничения на тези показатели. Всъщност реалността на тяхното разработване, използване и приемане показва проблем с валидността на конструкцията, при който включените бенчмаркове — поради тяхното инстанциране в конкретни данни, показатели и практика — не могат да обхванат нищо представително от твърденията за обща приложимост, които се правят за тях.



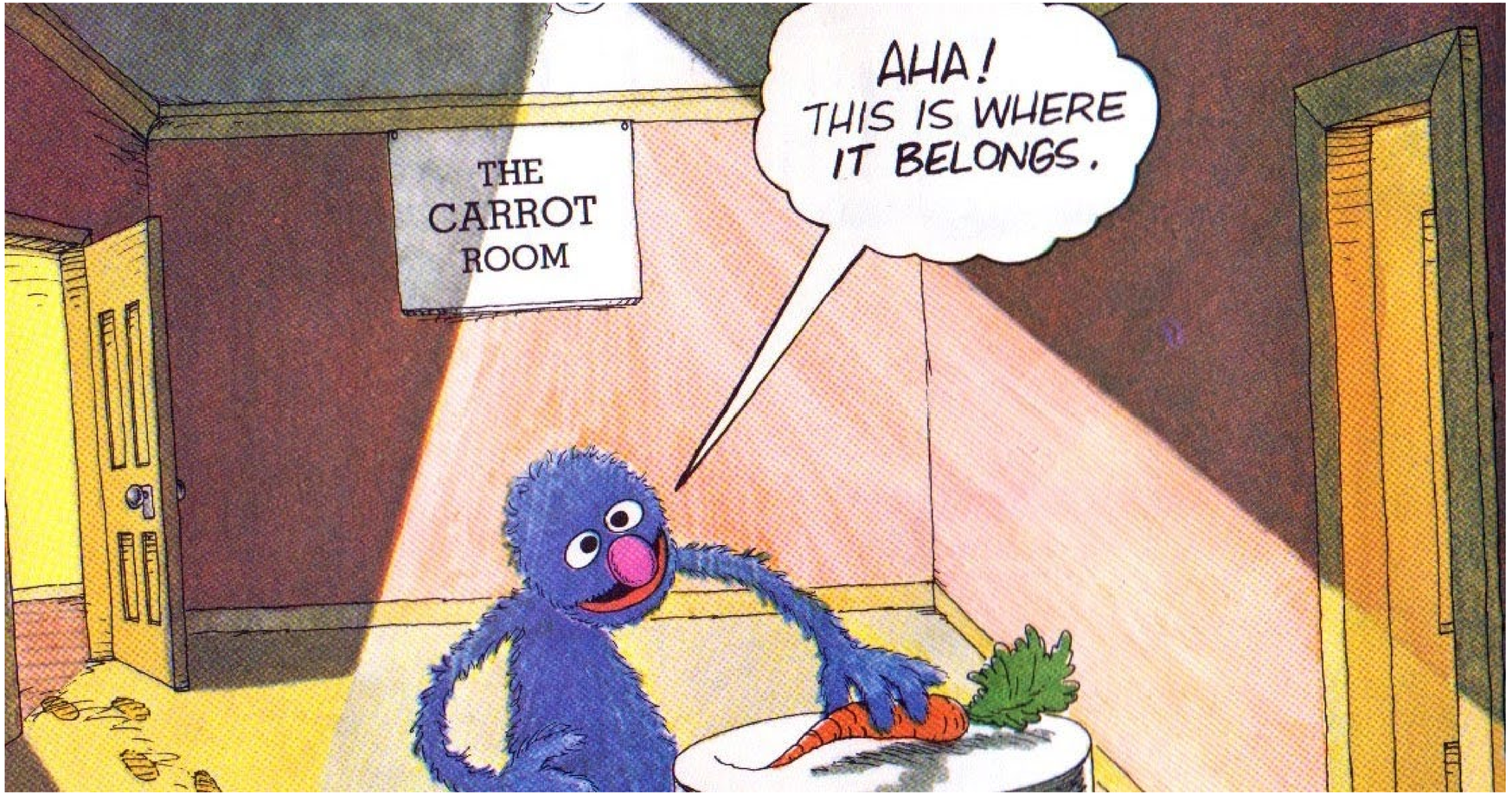
I.D. Raji, E.M. Bender, A. Paullada, E. Denton, A. Hanna, *AI and the Everything in the Whole Wide World Benchmark*, https://arxiv.org/abs/2111.15366

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Стаята с моркови

Аха, тук му е мястото.



Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Нещата, които виждате в стаята под водата.

Глъб! Това са все неща, които принадлежат под водата.
Освен мен! Махам се от тук.

Co-financed by the European Union
Connecting Europe Facility

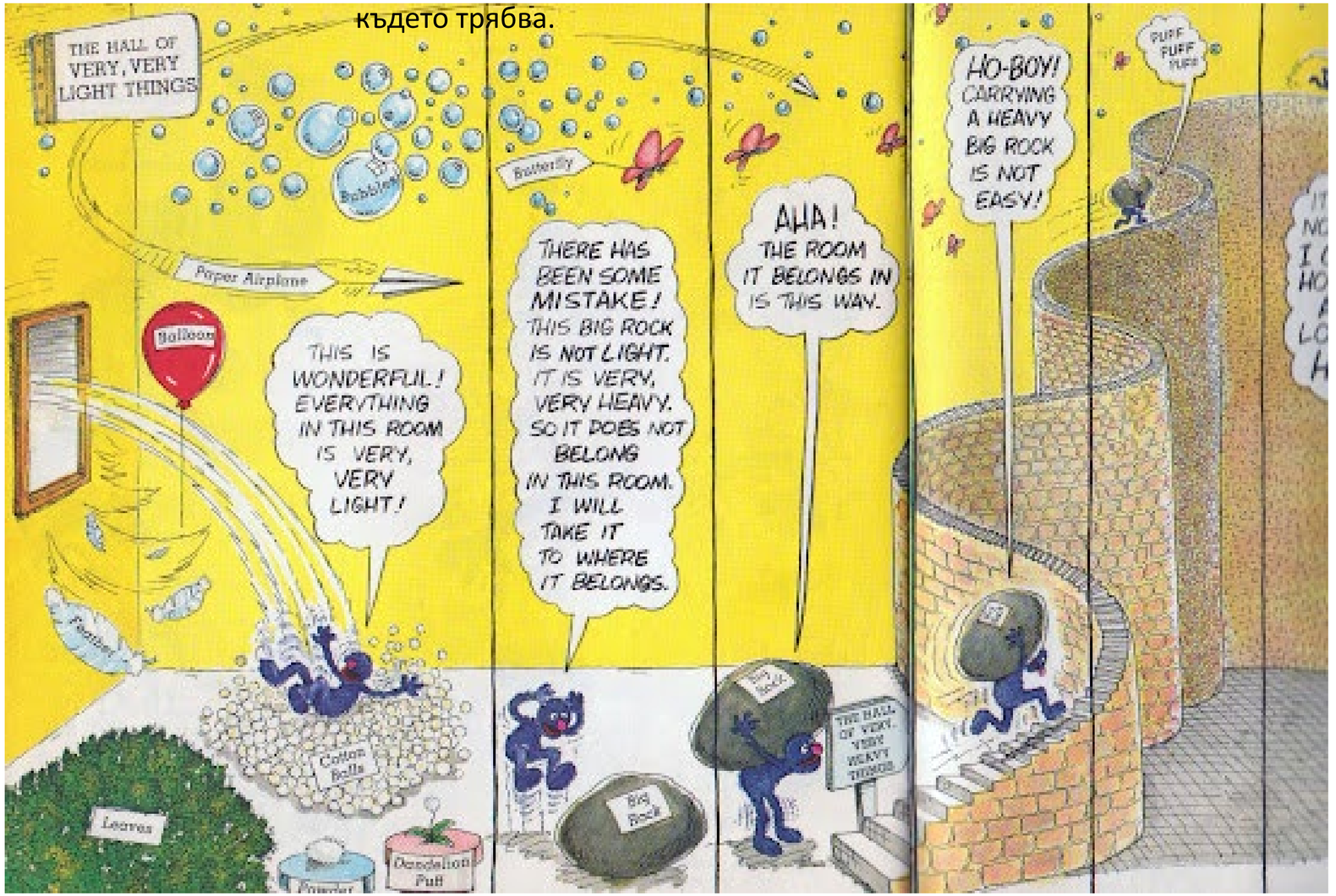
This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Залата на много много леките неща.

Това е чудесно. Всичко в тази стая е много много леко!

Тук има някаква грешка! Тази голяма скала не е лека. Тя е много тежка. Не ѝ е мястото в тази стая. Ще я занеса, където трябва.

Нейната стая е нататък. Не е лесно да носиш тежка скала.



Co-financed by the European Union
Connecting Europe Facility

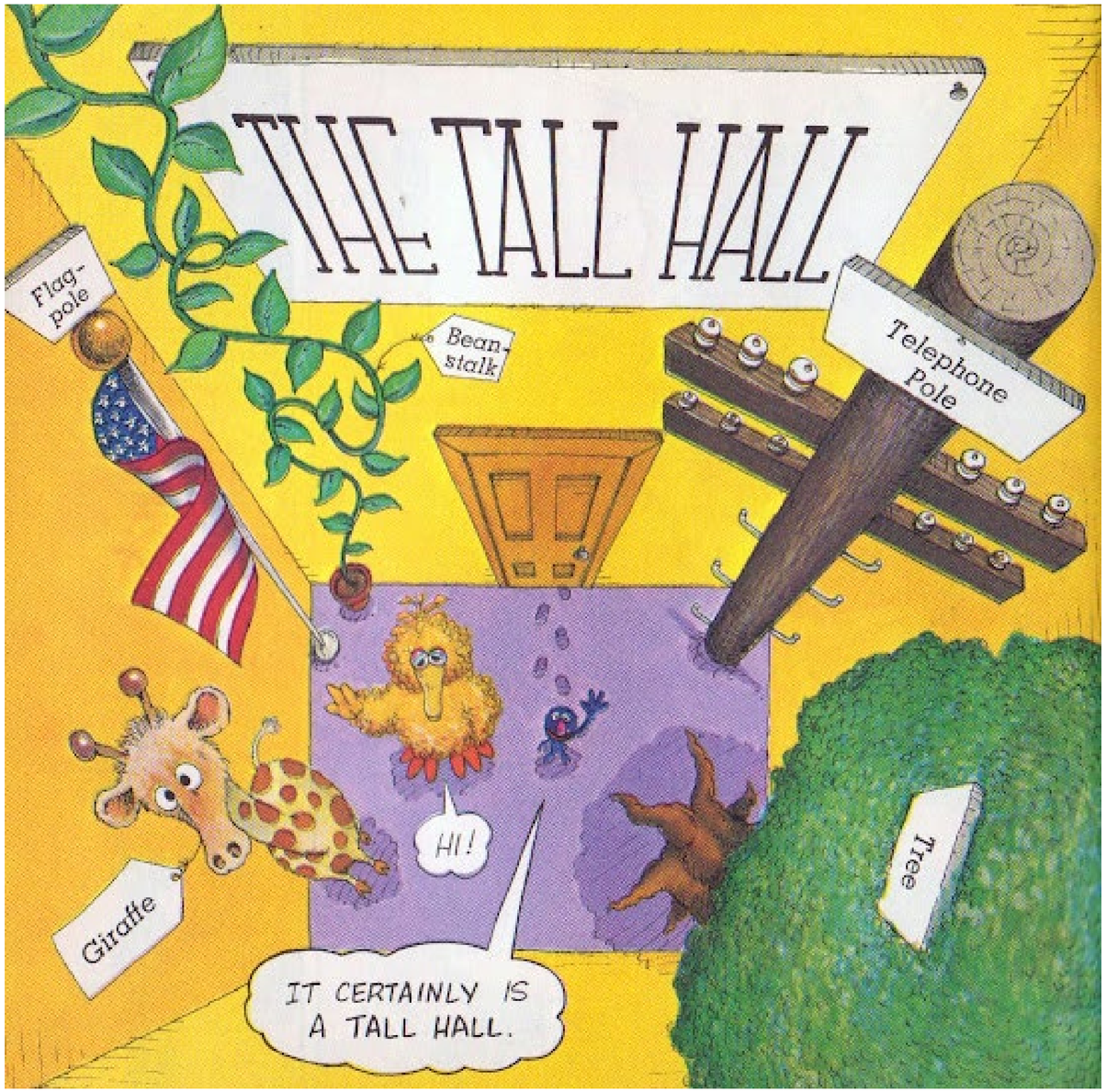
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Неща, които вдигат толкова много шум, че не можеш да си чуеш мислите.

Аааа! Не мога да си чуя мислите!

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

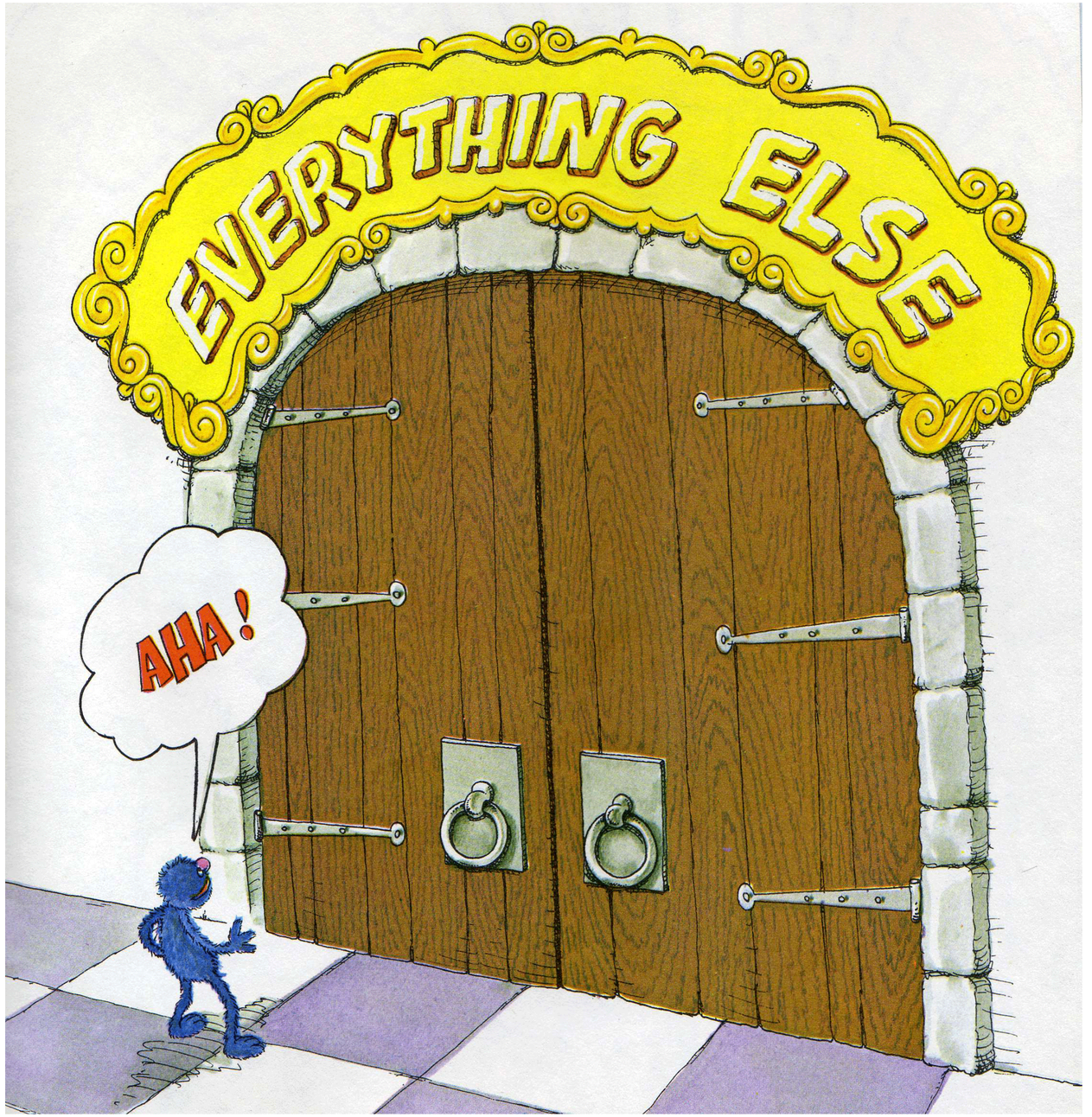


Високата стая!

Това със сигурност е висока стая.

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423



Всичко останало

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

ИИ и бенчмарк Всичко в целия свят (WWW)

“Ограничения на сравнителния анализ (бенчмарк) на общите възможности”

- “Въображаемият артефакт на „общия“ бенчмарк всъщност не съществува. Реалните данни са проектирани, субективни и ограничени по начини, които налагат различна рамка от тази на всяка претенция за общи знания или способности с общо предназначение. Всъщност представянето на който и да е единичен набор от данни по този начин в крайна сметка е опасно и измамно, което води до погрешни насоки относно дизайна и фокуса на задачите, недостатъчно отчитане на много пристрастия и субективни интерпретации, присъщи на данните, както и позволява, чрез фалшиви представяния на ефективността, потенциален модел злоупотреба”

- “бенчмаркингът е ограничен подход за оценка на общите възможности на модела”

I.D. Raji, E.M. Bender, A. Paullada, E. Denton, A. Hanna, *AI and the Everything in the Whole Wide World Benchmark*, https://arxiv.org/abs/2111.15366

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

ИИ и бенчмарк Всичко в целия свят (WWW)

“Ситуацията с Гроувър и твърденията в музея са очевидно нелепи – въпреки това в машинното обучение ние следваме абсолютно същите логически грешки, за да оправдаем повишаването на избран брой показатели, работещи като общи показатели за областта. Въпреки това, няма набор от данни, който да може да улови пълната сложност на детайлите на съществуването, по същия начин, по който не може да има музей, който да съдържа пълния каталог на всичко в целия свят. Не съществуват отворени, универсални и неутрални набори от данни и настоящите методи за сравнителен анализ не предлагат значими измервания на общи възможности.”

“**разбирането на езика разчита не само на езикова компетентност, но и на познания за света, разумни разсъждения и способност за моделиране на душевното състояние на събеседника, нито едно от които не може да бъде напълно тествано чрез задачи само с текст**, като GLUE. Няколко изследователи посочват необходимостта от установяване на ефективна физическа и социална основа като част от процеса на преминаване към стабилно и ефективно разбиране на естествения език, предупреждавайки срещу ученето само с текст като ограничен подход. Бендер и Колър допълнително споменават **тенденцията на изследователите на машинното обучение да тълкуват погрешно определени бенчмаркове като улавящи способността на модела да дешифрира смисъла на езика, като твърдят, че бенчмарковете трябва да бъдат конструирани внимателно, ако искат да покажат доказателства за „разбиране“, за разлика от простата способност за манипулиране на езикова форма в достатъчна степен, за да премине теста**.”

I.D. Raji, E.M. Bender, A. Paullada, E. Denton, A. Hanna, *AI and the Everything in the Whole Wide World Benchmark*, https://arxiv.org/abs/2111.15366

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

MAI4CAREU

Master programmes in Artificial Intelligence 4 Careers in Europe

Благодаря. Въпроси?

daniela.tafani@unibo.it

Co-financed by the European Union
Connecting Europe Facility

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423