

Не погоджуйтеся на імітації

ЕДВАРД ФЕЙЗЕР

6 ЛЮТОГО 2019

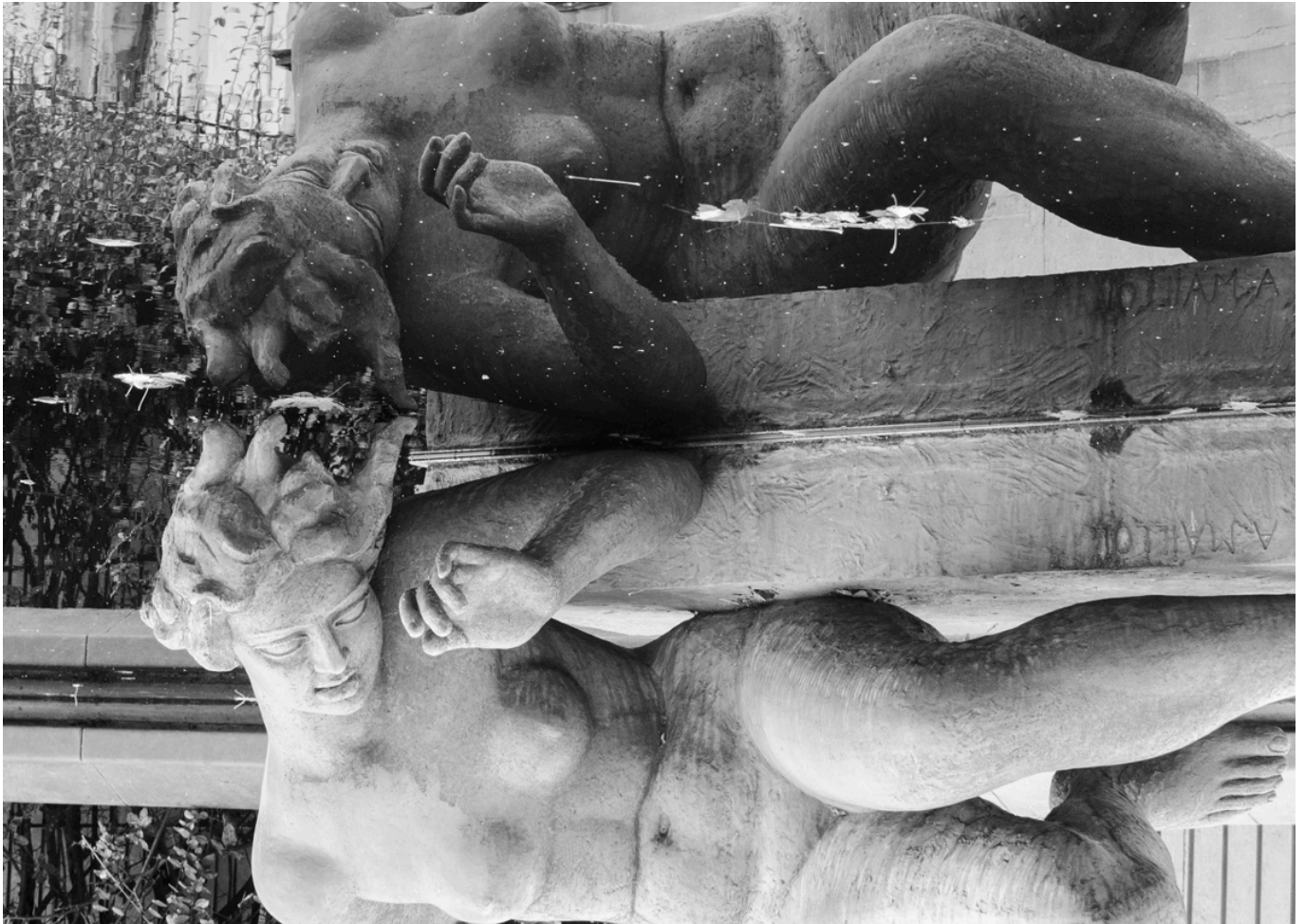


Фото: Daniele Imbriani

Алан Тюрінг нещодавно став зіркою кінофільму ¹, тож його класична праця «Обчислювальні машини і розум» ² здається доречною темою для роздумів. У цій статті Тюрінг описує знамениту «імітаційну гру», яку відтоді називають тестом Тюрінга. Її засаднича ідея така: припустімо, особа проводить допит, спілкуючись за допомогою клавіатури й монітора з двома учасниками – людиною і машиною – які перебувають у різних кімнатах. Завдання слідчого – встановити, хто є хто. Чи можна запрограмувати машину так, щоб із самої розмови не вдалося визначити, де людина, а де машина? Це питання Тюрінг запропонував як корисну заміну іншої

проблеми: чи здатні машини мислити? На його думку, відповідь «так» на перше запитання означатиме «так» і на друге.

Таке тлумачення проблеми вельми показове. Хоча в цій статті Тюрінг, власне, не стверджує, що машини здатні мислити або що комунікативна поведінка передбачає інтелект, таке враження, поза сумнівом, виникає (Джек Коупленд у нещодавній книжці, присвяченій Тюрінгові (с. 209)³, зазначає, що його висловлювання на цю тему не зовсім послідовні: у деяких випадках Тюрінг прямо відмовляється давати хоч якісь дефініції мислення, а в інших вказує, що вивчення роботи машин може допомогти нам з'ясувати, що ж таке мислення). Тюрінг пише, що запитання «Чи здатні машини мислити?» «занадто безглузде, аби заслуговувати на обговорення», що встановити, чи здатна машина впоратися з тестом Тюрінга означає розглянути «точнішу форму цього запитання» та що в тому разі, якщо машини розвинуться до рівня спроможності впоратися з тестом, «слововжиток і поінформована суспільна думка зміняться настільки, що можна буде говорити про машинне мислення, не очікуючи заперечень».

Чи може хтось вважати пірит золотом і чи пірит справді є золотом – це два різні запитання, тож я просто змінив би тему, а не відповів би на початкове запитання. Чим це відрізняється від Тюрінгової процедури?

Це дуже дивно. Припустімо, ви запитали мене, чи золото й пірит однакові, а я відповів, що запитання «занадто безглузде, аби заслуговувати на обговорення», і що було б «точніше» замислитися, чи можна обробити пірит так, аби його не вдалося відрізнити від золота. Якщо ж із цим упоратися, то «слововжиток і поінформована суспільна думка зміняться настільки, що можна буде говорити про пірит як про золото, не очікуючи заперечень». Звісно, це була б химерна відповідь. *Чи може хтось вважати пірит золотом і чи пірит справді є золотом – це два різні запитання, тож я просто змінив би тему, а не відповів би на початкове запитання. Чим це відрізняється від Тюрінгової процедури? І як саме запитання «Чи здатні машини мислити?» «безглуздіше», ніж запитання про тотожність піриту й золота?*

До речі, цього порівняння не вдасться відкинути на тому ґрунті, що, мовляв, експерт таки здатен відрізнити пірит від золота. У цьому сенсі порівняння *цілковито* виправдані, що імпліцитно визнавав і сам Тюрінг. Як вказує Коупленд (с. 211), він писав, що під час

«імітаційної гри» експерт, який розуміється на машинах, зможе за ледь помітними деталями встановити, де машина. Тож для Тюрінга умовою був *брак* у слідчого експертного досвіду. Він вважав, що важливіше те, чи зможе звичайна людина впізнати машину. Тож спроможність фахівця (на противагу звичайному спостерігачеві) визначити, чи є певний матеріал піритом, чи ні, аж ніяк не нівелює аналогічності мого прикладу.

Тож які Тюрінг – або будь-хто інший – мав підстави вважати, що такий тест підкаже, чи здатні машини мислити? Є принаймні три можливі відповіді, утім, жодна з них не вичерпна. Назвімо їх «схоластичною відповіддю», «верифікаціоністською відповіддю» і «сцієнтистською відповіддю» й розгляньмо кожну.

Те, що я називаю «схоластичною відповіддю», безперечно, відрізняється від того, що мав на увазі сам Тюрінг, проте це все-таки найперспективніший (хоч і зрештою безрезультатний) варіант захисту його процедури. Ідея така. Згадаймо основний принцип схоластичної метафізики: «*agere sequitur esse*» («дія йде за буттям» або «діяльність іде за існуванням»). Іншими словами, те, як річ *діє* або *поводиться*, віддзеркалює те, чим вона є. Прибічник тесту Тюрінга міг би вказати: якщо машина *діє*, як розумна істота, то вона мусить *бути* розумною істотою. Однак повноцінне використання мови – це парадигмально розумова діяльність (особливо для схоласта, який визначає інтелект із перспективи схоплення абстрактних понять, сформульованих у загальних термінах). Отже, міг би бути висновок, тест Тюрінга дозволяє безпомилково визначити, чи є інтелект.

Але зачекаймо. У схоластиці принцип «*agere sequitur esse*», безперечно, треба застосовувати в поєднанні з іншими засадничими метафізичними принципами. А в одному з них якраз і йдеться про різницю між *субстанційною* й *акцидентною формами*, яка виявляється в наявності або браку непіддатної спрощенням каузальності. Рослина здійснює фотосинтез, а кишеньковий годинник показує час, але каузальність для цих об'єктів не однакова. Фотосинтез – це *незалежний від спостерігача* факт про рослину, тоді як показування часу *не* є таким фактом про годинник. Металеві частини годинникового механізму не мають *природженої* властивості показувати час. Це функція, яку ми ніби ззовні нав'язали їм. Рослина ж має природжену властивість фотосинтезу. Отже, бути рослиною означає мати субстанційну форму, тобто бути справжньою субстанцією, тоді як бути кишеньковим годинником означає мати лише акцидентну форму, тобто не бути справжньою субстанцією. У цьому випадку справжньою субстанцією виявляються металеві запчастини, із яких складається годинник, а форма кишенькового годинника – це тільки акцидентна форма, яку ми нав'язали їм. (Я розглядав відмінність між субстанційною й акцидентною формами в різних текстах⁴⁵⁶; повний виклад див. у розділі 3 «Схоластичної метафізики»⁷)

Обчислювальна машина більше схожа на кишеньковий годинник, аніж на рослину. Вона запускає програми й долучається до розмови так само, як і годинник, який показує час. Себто, не маючи *природженої* властивості робити ці речі, вона виконує їх лише доти, доки ми нав'яжемо відповідні функції механізмам, із яких вона складається. (Саме тому, як вказує Сол Кріпке⁸, робота комп'ютерної програми не є незалежним від спостерігача фактом; і, як зазначають Карл Поппер і Джон Серль⁹, нема незалежного від спостерігача факту також у тому, чи певну річ узагалі можна вважати комп'ютером). Бути комп'ютером означає мати суто акцидентну, а не субстанційну форму.

Послуговуючись принципом «*agere sequitur esse*», треба спершу визначити, чи ми застосовуємо його до справжньої субстанції – тобто до об'єкта, який має субстанційну форму, а не лише акцидентну. Дослідивши металеві запчастини й виявивши, що вони показують час, було б безглуздо зробити висновок, що «*agere sequitur esse*», тож металеві запчастини можуть показувати час. Запчастини «показують час» тільки тому, що ми нав'язали їм цю функцію, інакше вони цього не робили б. Так само, жбурнувши камінь у повітря, було б смішно зробити висновок, що «*agere sequitur esse*», тому каміння може літати. Камінь «летить» лише завдяки тому, що я жбурнув його. Тож політ – це тільки акцидентна форма каменя. Послуговуючись принципом «*agere sequitur esse*», важливо встановити, що *природно* і *самостійно* робить річ, коли йдеться *суто про її властивості*, тобто виявити риси, які виростають із її субстанційної форми на противагу нав'язаній ззовні акцидентній формі.

З такої перспективи тест Тюрінга здається приреченою вигадкою. Спроможність машини мислити не дорівнює її здатності впоратися з тестом; ба більше, *якщо* вона це робить, то тільки тому, що її так запрограмували. Самі по собі металеві запчастини не показують часу, а каміння не літає. Залишені сам на сам машини не розмовляють одна з одною. І те, що ми здатні *змусити* їх розмовляти, не більше вказує на їхній інтелект, ніж жбурляння каменів або виготовлення годинників свідчить про спроможність каміння літати, а металевих запчастин – показувати час.

**Спроможність машини мислити не дорівнює її здатності
впоратися з тестом; ба більше, *якщо* вона це робить, то
тільки тому, що її так запрограмували. Самі по собі
металеві запчастини не показують часу, а каміння не
літає.**

Хоча схоластична відповідь була б (на мою думку, адже я схоласт) найкращим варіантом для Тюрінга, вона все-таки не дуже працює. Утім, Тюрінг схоластом аж ніяк не був. Можливо, він мав на увазі те, що я називаю «верифікаціоністською відповіддю»? Тут би йшлося про таке: верифікаціонізм вважає, що значення висловлювання визначає метод його верифікації. Однак ми не можемо зануритися в чужий розум – як машинний, так і людський. Тому, можна було б міркувати, єдиний спосіб перевірити, чи річ має інтелект, – це встановити, чи вона *поводиться* розумно, а золотий стандарт розумної поведінки – розумна розмова. Тому адекватно інтерпретувати запитання «Чи здатні машини мислити?» можна, тільки перефразувавши його в «Чи здатні машини розумно поводитися?». Тест Тюрінга саме це й намагається встановити, тож, якщо машина успішно з ним впорається, більше нічого й не буде потрібно, щоб засвідчити реальну наявність у неї інтелекту. Справді, можна було би припустити, інтелект не може *бути* нічим іншим, тільки здатністю скласти тест Тюрінга.

Коли писав Тюрінг, верифікаціонізм, безумовно, був на часі. На ньому ґрунтується ідея «філософського біхевіоризму» про те, що мати розум – це «не що інше, як» виявляти певні поведінкові моделі та схильності. Однак є низка проблем, пов'язаних із верифікаціонізмом, і не остання з них полягає в тому, що він сам собі суперечить. Принцип верифікації неможливо верифікувати, тож він безглуздий за власними стандартами. Якби він був послідовний, то не досяг би навіть рівня хибності. Не дивно, що ніхто його більше не відстоює – принаймні в найпрямолінійнішій формі.

Але Тюрінг у жодному разі не звертається до верифікаціонізму, і, мабуть, тут не про цю концепцію йдеться. Гадаю, Тюрінг схилився – нехай і не висловлюючи цього прямо – до «сцієнтистської відповіді» на запитання, чому можна вважати, що «імітаційна гра» підкаже, чи здатні машини мислити. Думка Тюрінга, як на мене, полягала в тому, що єдиний спосіб вивчати інтелект із *наукової* точки зору – це запитувати, яка ж має бути система, щоб упоратися з тестом. Тож саме на цьому питанні, вважав він, треба зосередитися. Звернімо увагу, що це не те (чи не має бути те), на чому наполягає верифікаціонізм. Усупереч своїм словам про «безглуздість», Тюрінг не стверджував, що *зовсім* безглуздо запитувати, чи може машина скласти тест, але насправді не мислити. Він міг лише вказати, що раз нема наукових інструментів для розгляду цього конкретного питання, то нема й сенсу ним перейматися. А тому варто зосередитися на тому, що можна пояснити методами емпіричної науки.

Якщо Тюрінг схилиється до цього, то він, по суті, робить те саме, що й Лоуренс Краусс, коли вдає, ніби відповідає на знамените питання, чому взагалі існує щось, а не ніщо. Однак те, що робить Краусс, – це чистісіньке шахрайство, якя вказував у кількох текстах.¹⁰¹¹¹²¹³ Спершу він удає, ніби збирається пояснити, чому існує щось, а не ніщо, але згодом змінює тему й обговорює натомість, як усесвіт у його нинішньому стані виник із порожнього простору й

законів фізики – які, погодьмося, аж ніяк не *ніщо*. Цей фарс, на його думку, можна виправдати тим, що фізика може нам щось відповісти на оте друге запитання, але їй нічого сказати про те, чому існує щось (зокрема й фундаментальні закони фізики), а не ніщо. З погляду Краусса, потрібно зосередитися на питаннях, на які, як йому здається, він здатен відповісти, а не на тих, які ми спершу поставили.

Це саме та помилка, якої припускається людина напідпитку, шукаючи під лампою загублені ключі від автомобіля лише на тій підставі, що це єдине місце, де достатньо світла, аби їх побачити. Те, що світло є саме тут, аж ніяк не означає, що ключі теж тут будуть; ба більше, це не означає, що є сенс знов і знов шукати ключі під лампою, коли їх уже кілька разів там не виявилось, чи що немає сенсу шукати в іншому місці, чи що треба шукати не ключі, а щось інше. Аналогічно, потужність методів фізики не гарантує їхньої спроможності відповісти, чому взагалі існує щось, а не ніщо. Це також не означає, що потрібно замінити це питання іншим, із яким фізичні методи спроможні впоратися, або що нема підстав дошукуватися інших методів дослідження проблеми. Припустити за Крауссом, що питання просто *мусить* бути придатним для фізичного дослідження, означає схибити на тому, що Едвін Бертт¹⁴ визначив як помилку «вилучення метафізики з царини методів», себто підпорядковувати реальність уподобаному методу, а не шукати метод, відповідний реальності.

**Потужність методів фізики не гарантує їхньої
спроможності відповісти, чому взагалі існує щось, а не
ніщо. Це також не означає, що потрібно замінити це
питання іншим, із яким фізичні методи спроможні
впоратися.**

Здається, Тюрінг теж на це хибує. Замість того, щоб спершу дати дефініцію мислення, а потім дошукуватися придатних методів для вивчення цього явища, він звертається тільки до тих пов'язаних із мисленням феноменів, які можна дослідити методами емпіричних наук, і доходить висновку, що *лише вони* вартують вивчення. Помилковість цієї логіки очевидна. Опис «мислення» як речі, яку машина виявить у собі, упоравшись із тестом Тюрінга, нагадує опис «ключів» як речі, яку можна знайти під певною лампою.

Загалом нема (як я не раз казав) жодної підстави приймати сцієнтизм і є переконливі причини

відмовитися від нього. Із ним пов'язано щонайменше п'ять проблем. По-перше, сцієнтистські тези зазвичай або неспростовні, або тривіально слухні. По-друге, природнича наука принципово не здатна запропонувати повний опис бодай тільки фізичного світу. По-третє, вона насправді не може вичерпно пояснити феномени, які описує. По-четверте, головний аргумент сцієнтизму – покликання на прогнозованій технологічні успіхи науки – помилковий. Нарешті, по-п'яте, поширене припущення, що єдиною альтернативою природничій науці є сумнівний метод «концептуального аналізу», також хибне. (Див. розділ 0 «Схоластичної метафізики»¹⁵, де детально розглянуто кожну з цих тез). Тож «сцієнтистська відповідь» теж не задовільна.

Поза сумнівом, Тюрінг був блискучим ученим, і всі ми, хто користується комп'ютерами й любить їх, перед ним у боргу. Однак його набіг на філософію виявився, на мою думку, не позитивним вкладом, а лише цікавою й повчальною помилкою.

Текст із авторського блогу. *Переклала Галина Помилуйко.*

ЕДВАРД ФЕЙЗЕР

6 ЛЮТОГО 2019

Warning: count(): Parameter must be an array or an object that implements Countable in </home/kairosbo/verbum.com.ua/www/wp-includes/post-template.php> on line **284**

Примітки

1. Ідеться про фільм 2014 року «Гра в імітацію» режисера Мортена Тільдума.
2. Turing, A. M. (1950). Computing Machinery and Intelligence. *Mind*, 59(236), 433–460.
3. Copeland, B. J. (2012). *Turing: Pioneer of the Information Age*. Oxford: Oxford University Press.
4. Feser, E. (2011). *Nature versus art*. Retrieved from <http://edwardfeser.blogspot.com/2011/04/nature-versus-art.html>
5. Feser, E. (2011). *Reply to Torley and Cudworth*. Retrieved from <http://edwardfeser.blogspot.com/2011/05/reply-to-torley-and-cudworth.html>

6. Feser, E. (2014). *Welcome to the machine*. Retrieved from <http://edwardfeser.blogspot.com/2014/04/welcome-to-machine.html>
7. Feser, E. (2014). *Scholastic Metaphysics: A Contemporary Introduction*. Frankfurt: Ontos Verlag.
8. Feser, E. (2012). *Kripke contra computationalism*. Retrieved from <http://edwardfeser.blogspot.com/2012/05/kripke-contra-computationalism.html/>
9. Feser, E. (2012). *Popper contra computationalism*. Retrieved from <http://edwardfeser.blogspot.com/2012/02/popper-contra-computationalism.html>
10. Feser, E. (2012). *Not Understanding Nothing*. Retrieved from <https://www.firstthings.com/article/2012/06/not-understanding-nothing>
11. Feser, E. (2013). *Forgetting nothing, learning nothing*. Retrieved from <http://edwardfeser.blogspot.com/2013/02/forgetting-nothing-learning-nothing.html>
12. Feser, E. (2013). *A gigantic book royalty check from nothing*. Retrieved from <http://edwardfeser.blogspot.com/2013/08/a-gigantic-book-royalty-check-from.html>
13. Feser, E. (2012). *Steng operation*. Retrieved from <http://edwardfeser.blogspot.com/2012/04/steng-operation.html>
14. Burt, E. A. (2003). *The Metaphysical Foundations of Modern Science*. New York: Dover Publications.
15. Feser, E. (2014). *Scholastic Metaphysics: A Contemporary Introduction*. Frankfurt: Ontos Verlag.