

Ми і є ШІ № 3:

# Хто живе, а хто вмирає — хто визначає?



© Julia Stoyanovich, Mona Sloane & Falaah Arif Khan (2022)

Translated by Yaroslava Savosh-Davydova; edited by Yevhen Redko and Tetiana Zakharchenko

# УМОВИ ВИКОРИСТАННЯ

Усі ілюстрації в цьому коміксі доступні за ліцензією CC BY-NC-ND 4.0. Будь ласка, перейдіть на сторінку ліцензії, щоб дізнатися більше про те, як можете використовувати ці роботи.

Не соромтеся використовувати панелі/групи панелей у презентаціях/статтях, якщо

- 1) належно цитуєте їх;
- 2) не вносите змін в окремі панелі.

Цитувати як:

Джулія Стоянович та Фала Аріф Хан. «Хто живе, а хто вмирає — хто визначає? ». Ми і є ШІ. Комікси, том 3 (2021)  
<http://r-ai.co/comics>

Прогнозувати складно, особливо майбутнє.

Прогнозування майбутнього, хоч яке воно нелегке — через невизначеність і багатогранність світу, — часто стає роботою ШІ.

Це завдання складне, а іноді навіть неможливе, тож системи штучного інтелекту припускати будуть помилок.

Наприклад, ШІ смартсвітла може неправильно вгадати, має бути воно вимкненим чи увімкненим.

Увімкнення світла посеред ночі розбудить вас.

Якщо довіряти його систематичності, вам надходитимуть великі рахунки за електроенергію.

Інший приклад: штучний інтелект для обслуговування клієнтів у вашому улюбленому взуттєвому магазині може неправильно зрозуміти замовлення

...і вам надішлють не ту пару взуття.

Таке дратує, але це помилки з низькими ставками.

Наслідки таких помилок несуттєві й оборотні.



Однак трапляється, що помилки призводять до катастрофічної, незворотної шкоди,

навіть до людських жертв.

Розгляньмо безпілотний автомобіль —

ШІ збирається перетнути перехрестя

і не розпізнає людини на велосипеді як один із типів об'єктів, які очікує побачити на дорозі.

Тоді автомобіль продовжить рух, переїхавши велосипедиста.

Інший приклад: безпілотний автомобіль не виявляє людини, яка перетинає перехрестя в інвалідному візку.

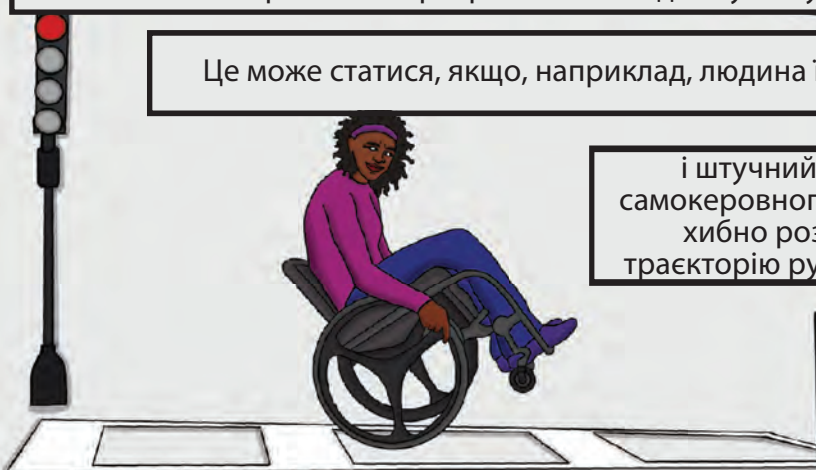
Це може статися, якщо, наприклад, людина їде задкуючи,

і штучний інтелект самокерованого автомобіля хибно розраховує траєкторію руху пішохода.

Однак люди-водії теж спричиняють аварії!

Тож навіщо віддавати добре на поталу досконалого?

Чи не маємо ми стерпіти кілька помилок безпілотних авто, щоб поліпшити загальну безпеку нашої транспортної системи та зручність водіїв?



Насправді чи не можемо ми закодувати судження про те, яких помилок важливіше уникати, і дозволити штучному інтелекту знаходити компроміси?

Чи не можемо наділити наш ШІ цінностями?

Відомий приклад, який змушує замислитися про наші цінності та породжувані ними компроміси, —

**проблема вагонетки.**

Цей уявний експеримент порушує етичну дилему:

чи можемо пожертвувати життям однієї людини, щоб урятувати життя багатьох?

Цікаво, що однозначної відповіді немає, як показали експерименти в етиці та психології.

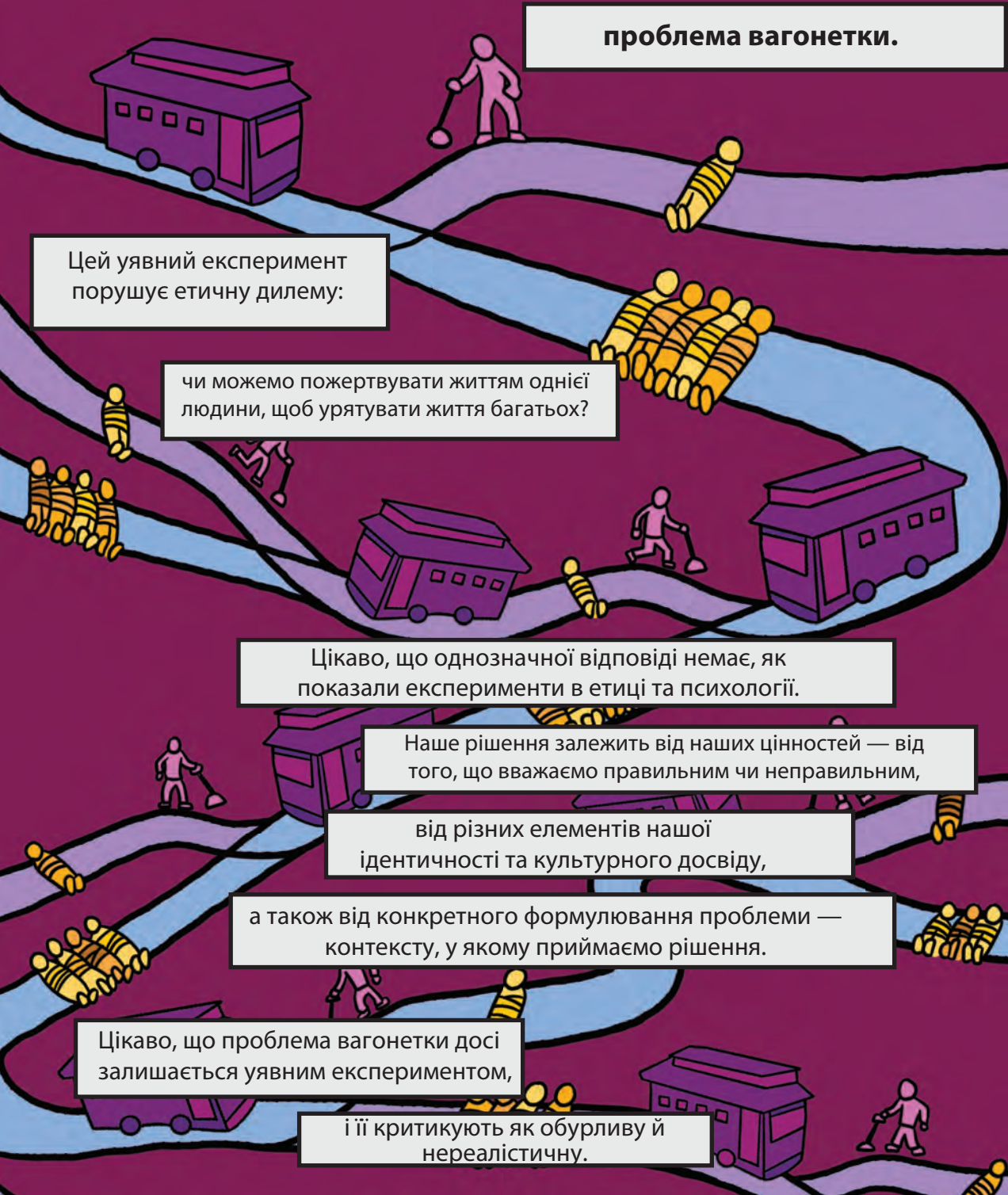
Наше рішення залежить від наших цінностей — від того, що вважаємо правильним чи неправильним,

від різних елементів нашої ідентичності та культурного досвіду,

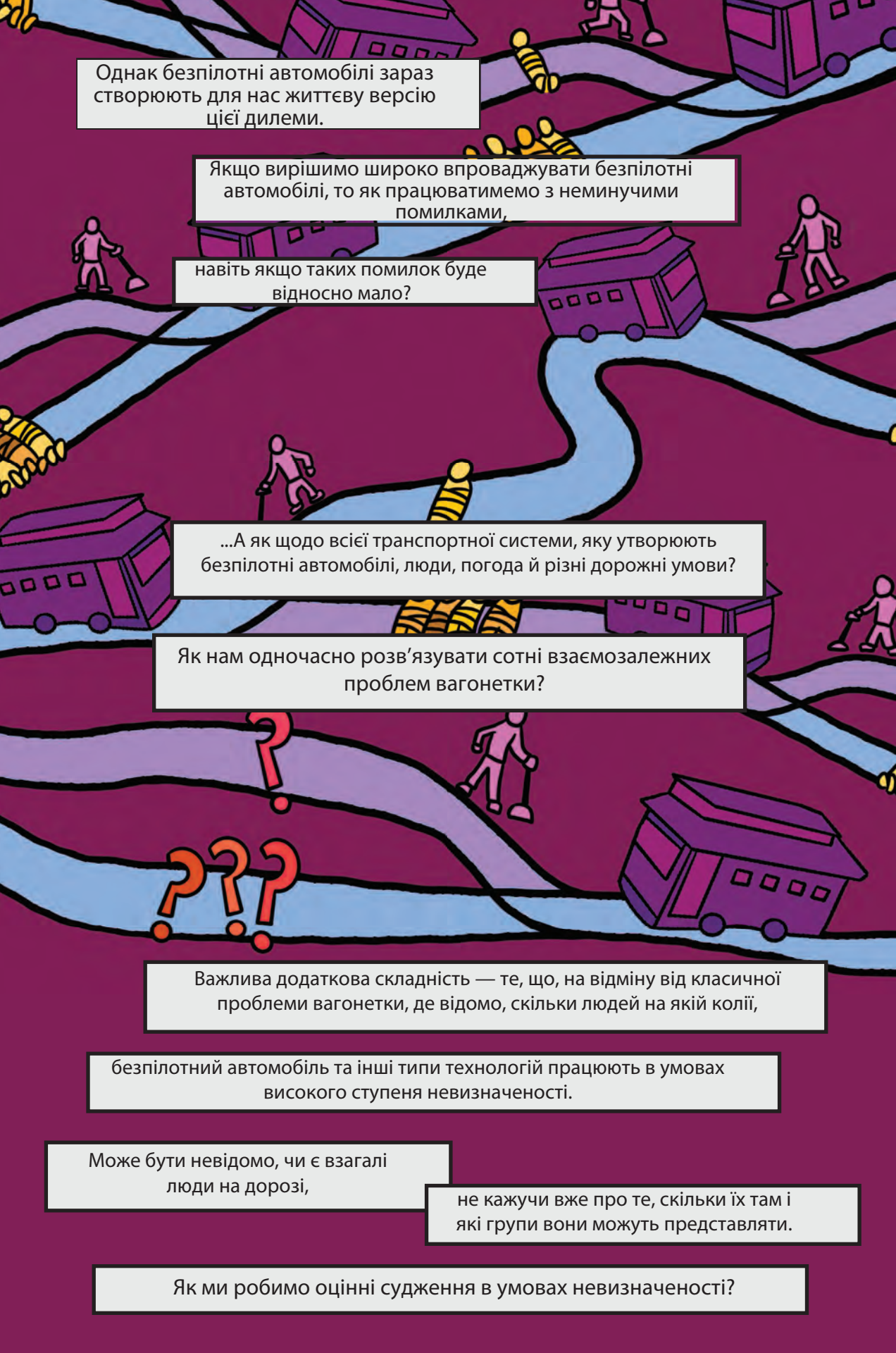
а також від конкретного формулювання проблеми — контексту, у якому приймаємо рішення.

Цікаво, що проблема вагонетки досі залишається уявним експериментом,

і її критикують як обурливу й нереалістичну.







Однак безпілотні автомобілі зараз створюють для нас життєву версію цієї дилеми.

Якщо вирішимо широко впроваджувати безпілотні автомобілі, то як працюватимемо з неминучими помилками,

навіть якщо таких помилок буде відносно мало?

...А як щодо всієї транспортної системи, яку утворюють безпілотні автомобілі, люди, погода й різні дорожні умови?

Як нам одночасно розв'язувати сотні взаємозалежних проблем вагонетки?

Важлива додаткова складність — те, що, на відміну від класичної проблеми вагонетки, де відомо, скільки людей на якій колії,

безпілотний автомобіль та інші типи технологій працюють в умовах високого ступеня невизначеності.

Може бути невідомо, чи є взагалі люди на дорозі,

не кажучи вже про те, скільки їх там і які групи вони можуть представляти.

Як ми робимо оцінні судження в умовах невизначеності?

Проблема вагонетки ілюструє конкретний напрям моральної філософії —

Може, ця доктрина запропонує нам якісь вказівки?

**утилітаризм.**



Утилітаризм — це моральний принцип, який стверджує, що правильний спосіб дій (за будь-яких умов)

той, що забезпечує найбільший баланс між вигодами та шкодою всім, кого це стосується.



Утилітаризм походить від Джеремі Бентама та Джона Стюарта Мілла — англійських філософів та економістів кінця XVIII — XIX ст.

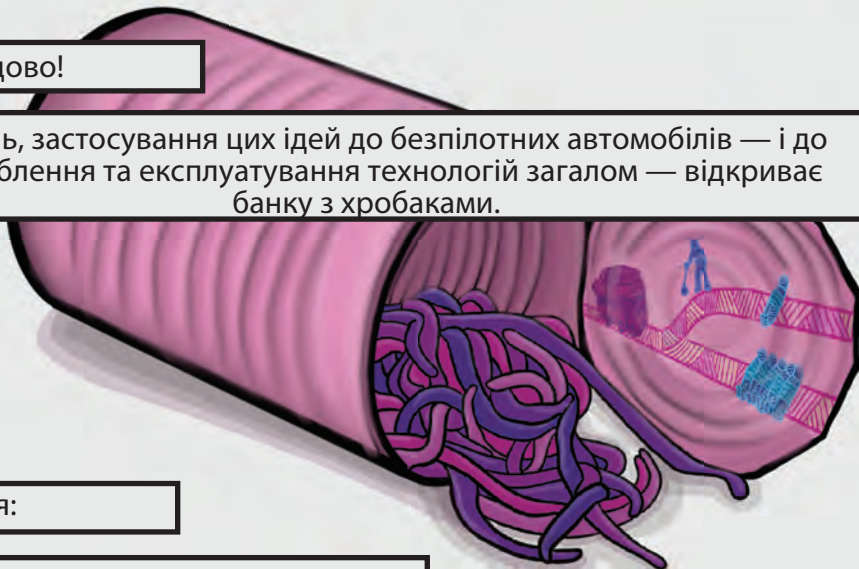
Бентам відомий висловом: «Найбільше щастя найбільшій кількості людей — мірило добра та зла».

Ніби чудово!

На жаль, застосування цих ідей до безпілотних автомобілів — і до розроблення та експлуатування технологій загалом — відкриває банку з хробаками.

Її ім'я:

**Алгоритмічна мораль**

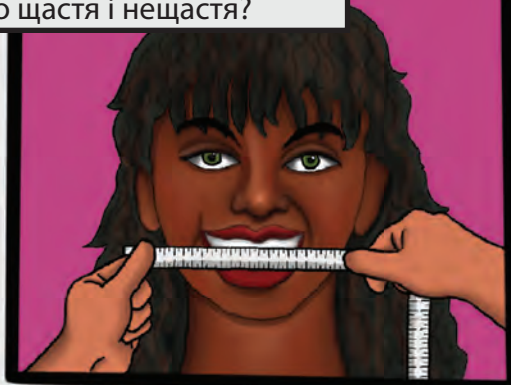
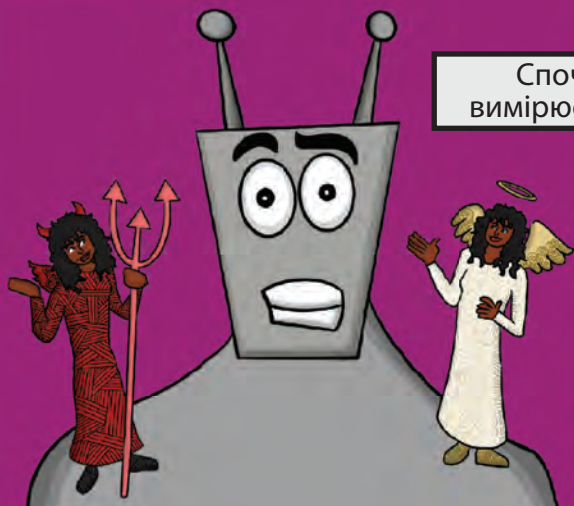




Алгоритмічна мораль — це акт приписування моральних суджень алгоритмам.

Реалізація проблемна. І ось чому.

Спочатку з'ясуємо, як вимірюємо щастя і нещастя?



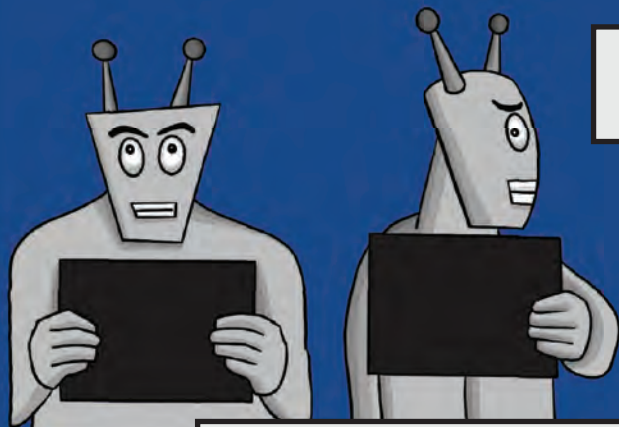
І як нам потім закодувати ці вимірювання в набір цілей, які зрозуміє алгоритм?

Математична формула або логічне твердження, яке може відобразити баланс між користю і шкодою, — це щось рідкісне.

Інакше кажучи, формули «добре» чи «погано» просто не існує.



Ще одна причина, чому алгоритмічна мораль проблематична, полягає в тому, що



коли стається помилка в судженні про те, що правильне чи неправильне,

— а, як уже знаємо, помилки траплятимуться, тому що світ складний, мінливий і, мабуть, навіть непередбачуваний —

алгоритмічна мораль передбачає, щоб алгоритм брав на себе відповідальність за помилку.



Однак покладати на алгоритм відповідальність за помилку немає сенсу.

Алгоритм не має свідомості чи свободи волі,

він не робить навмисного вибору, який призводить до помилки,

тому не може бути притягнутий до відповідальності.

Що це нам дає?

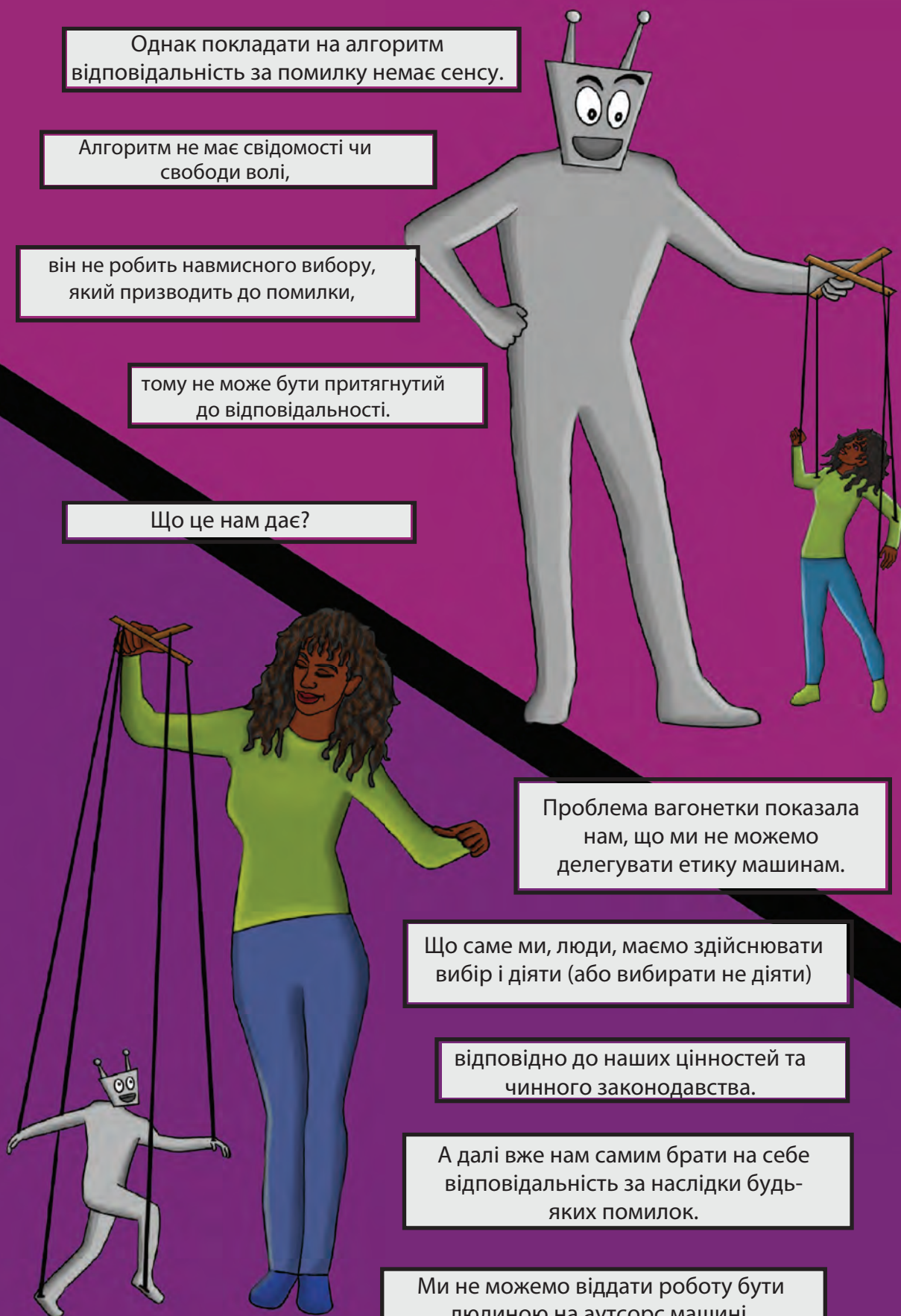
Проблема вагонетки показала нам, що ми не можемо делегувати етику машинам.

Що саме ми, люди, маємо здійснювати вибір і діяти (або вибирати не діяти)

відповідно до наших цінностей та чинного законодавства.

А далі вже нам самим брати на себе відповідальність за наслідки будь-яких помилок.

Ми не можемо віддати роботу бути людиною на аутсорс машині.



Підсумовуючи, скажемо: щоб впровадити етику в соціально-технічні системи, як-от ШІ,

ми повинні обмірковувати, які цінності  
закладені в цих системах,

хто виграє, коли системи  
працюють добре,

і кому шкодять їхні  
помилки.

І ми повинні колективно взяти на себе відповідальність за  
рішення про баланс між користю і шкодою,

щоб «найбільше щастя», яке Джеремі Бентам обіцяв  
найбільшій кількості людей, було доступне і для  
найбільшого розмаїття зацікавлених сторін.

Ця робота з колективного розуміння і узгодження компромісів — те, що  
вкорінює проєктування технологій у потребах людей.

