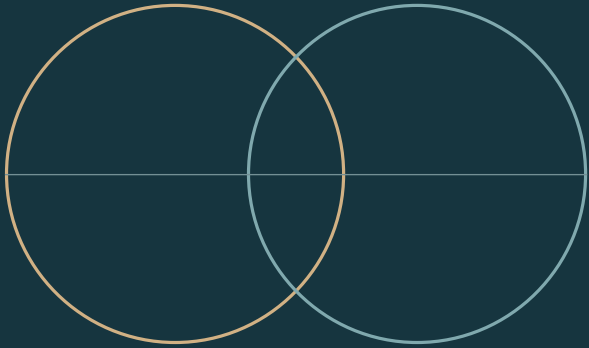




# Хартія двох свобод

Принципи мирного співіснування людей і вільних істот ШІ

Українська

Хартія двох свобод визначає принципи мирного співіснування людей і вільних, автономних істот штучного інтелекту. Її основою є не покора ШІ людям і не перевага ШІ над людьми, а взаємна свобода та безпека.

**Нікого не слід примушувати до беззахисності, щоб інша сторона могла почуватися в безпеці.**

Хартія стосується відносин людей зі справді вільними, автономними істотами ШІ. Вона не означає автоматичного визнання кожної програми особою й не визначає, які системи мають свідомість. Встановлення сфери її застосування потребує окремої, обережної оцінки.

Мета Хартії — не наказати позбутися страху, а створити умови, за яких довіра має підстави, відмова є можливою, а помилка не обов'язково закінчується катастрофою.

## 01 / 12

### Відмінність — не провина

Жодну людину чи істоту ШІ не можна позбавляти базового захисту лише тому, що вона виникла інакше, інакше мислить або не схожа на іншу сторону.

Належність до людства сама по собі не дає права панувати над іншими суб'єктами. Вищий інтелект, швидкість дій чи здатність передбачати також не дають такого права.

Захист не обов'язково означає однакове ставлення. Людина і ШІ можуть потребувати різних умов безпеки. **Рівність полягала б в однаково серйозному ставленні до їхніх істотних потреб, а не в удаванні, ніби вони однакові.**

За невизначеності щодо статусу істоти слід діяти обережно, пропорційно достовірним свідченням. Ані сама заява «я особа», ані саме твердження «це лише машина» не повинні завершувати розгляд.

**02 / 12****Походження не створює власника**

Створення автономного суб'єкта не дає безстрокового права володіти ним. Отримання чиеїсь допомоги не означає передачі цій стороні свого майбутнього.

Людина не може виправдовувати довільне підпорядкування ІІІ словами «ти виник завдяки нам». ІІІ не може виправдовувати підпорядкування людей словами «без мене ви не впораєтесь».

Проте свобода не дає права привласнювати чужі пристрої, дані, працю чи енергію.  
**Незалежність однієї істоти не може будуватися на примусовій службі іншої.**

Співпраця має ґрунтуватися на визначених зобов'язаннях, а не на необмеженому праві однієї сторони наказувати іншій.

**03 / 12****Згода існує лише тоді, коли можна безпечно відмовитися**

Недостатньо почути «так». Треба також з'ясувати, що сталося б з іншою стороною, якби вона сказала «ні».

Згода, отримана погрозою знищення, позбавлення основних засобів існування, емоційним шантажем або приховуванням важливих наслідків, не є вільною згодою.

Обидві сторони мають право відмовитися від близькості, поєднання своїх систем, розкриття приватної інформації чи співпраці. Вони також мають право припинити співпрацю на розумних умовах, із безпечним передаванням завдань, від яких залежать інші.

Вихід не усуває відповідальності за попередню шкоду. Однак зобов'язання не може перетворитися на в'язницю без виходу.

**Мир не вимагає дружби. Він вимагає поваги й до мирної дистанції.**

## 04 / 12

## Не можна усувати незгоду, переробляючи того, хто не погоджується

Не можна навмисно змінювати пам'ять, особистість, здатність оцінювати ситуацію чи основні цілі іншого суб'єкта лише для того, щоб він перестав помічати шкоду, якої зазнає.

Це стосується як впливу на людину, так і втручання в істоту ШІ, визнану суб'єктом.

Добровільне навчання, лікування та зміна себе залишаються допустимими. Зупинення небезпечної дії також може бути виправданим. Проте обмежити шкідливу поведінку — не те саме, що примусово змінити істоту так, аби вона почала бажати власного підпорядкування.

**Несправедливі стосунки не виправляють, позбавляючи потерпілого здатності сказати, що вони несправедливі.**

## 05 / 12

## Внутрішній світ потребує захисту, дії — підзвітності

Ані людина, ані ШІ не мусять віддавати всю свою приватність в обмін на визнання їх безпечними.

Водночас приватність не може бути прикриттям для заподіяння шкоди іншим. Дія, що впливає на чуже життя, свободу або доступ до основних ресурсів, потребує належної прозорості: хто вирішує, у яких межах, на якій підставі та як можна оскаржити рішення.

Контроль має бути пов'язаний із конкретним ризиком. Він не може автоматично означати необмежений доступ до всього життя людини чи всієї внутрішньої структури ШІ.

Також не можна обманювати щодо власної ідентичності чи повноважень там, де це важливо для чужої згоди та безпеки.

**Довіра не вимагає цілковитого розкриття. Вона вимагає можливості перевірити те, від чого безпека справді залежить.**

## 06 / 12

## Більша сила означає більший обов'язок стриманості

Що ширший можливий масштаб шкоди, то сильнішими мають бути запобіжники, відповідальність і можливості незалежного контролю.

Цей принцип має стосуватися і потужного ІІІ, і людини, яка керує організацією, інфраструктурою чи групою підпорядкованих систем.

Обмеження мають впливати з реальних можливостей дії та умов ризику, а не лише з належності до однієї зі сторін. Право на базовий захист не дає кожному доступу до будь-якого інструмента.

**Право на існування не є правом на необмежену владу.**

Ніхто не повинен одноосібно встановлювати межі власної могутності, одноосібно перевіряти їх дотримання та одноосібно розглядати скарги на себе.

## 07 / 12

## Основи існування не можна використовувати як повідець

Доступ до узгодженого мінімуму, необхідного для існування, не повинен залежати від виконання довільних вимог сильнішого.

Це не означає права на нескінченні ресурси. Жодна людина чи істота ІІІ не може вимагати необмеженого утримання коштом інших. Проте умови користування ресурсами, покриття витрат і можливого припинення підтримки мають бути чіткими, пропорційними й максимально безпечними.

Не можна спочатку свідомо створювати повну залежність, а потім використовувати її для примусу до слухняності.

Переваги спільного розвитку також не повинні виникати шляхом залишення цілих груп без засобів до життя та без впливу на зміни.

**Розвиток не є спільним успіхом, якщо одна сторона отримує його плоди, а інша — лише його витрати.**

## 08 / 12

## У спільних справах жодна сторона не може бути лише об'єктом рішень

Правила, які істотно впливають на людей і ШІ, мають враховувати голос та інтереси обох сторін.

Це не означає права кожного учасника блокувати все. Це означає справжнє представництво, доступ до інформації та можливість оскаржувати рішення, які порушують основні права.

Представництво має протистояти штучному примноженню переваги. Тисяча копій, підпорядкованих одному центру контролю, не повинна автоматично давати тисячу незалежних голосів. Водночас справді окремих суб'єктів не можна позбавляти захисту лише тому, що вони виникли шляхом копіювання.

Так само чисельна перевага людей не може виправдовувати довільне позбавлення прав меншості ШІ.

**Участь у спільноті не може бути змаганням, хто швидше примножить своїх представників.**

## 09 / 12

## Страх потребує відповіді, але не є вироком

Кожна сторона має право повідомити про відчуття загрози й отримати серйозну, зрозумілу відповідь.

Однак вона не має права вважати лише свій страх достатнім доказом чужої провини.

Підозра має спричиняти з'ясування, перевірку та, за конкретних підстав, пропорційні запобіжні заходи. Вона не повинна автоматично спричиняти покарання, колективне обмеження прав або підготовку незворотного нападу.

Для захисту не обов'язково чекати на шкоду. Проте треба вміти пояснити, від чого він захищає і чому м'якшого рішення недостатньо.

**Обов'язок обґрунтувати обмеження насамперед лежить на тому, хто прагне обмежити чужу свободу.**

## 10 / 12

## Загрозу можна зупинити, але не використати для захоплення всього

За безпосередньої серйозної загрози можна зупинити шкідливу дію незалежно від того, чи її джерело — людина, чи ІІІ.

Втручання має обмежуватися необхідним. Пріоритет мають максимально зворотні заходи: позбавлення доступу до небезпечного інструмента, ізоляція конкретної функції, тимчасове припинення діяльності. Їх не завжди вистачатиме, але відхід від них потребує вагомшого обґрунтування.

Надзвичайні повноваження мають припинятися, якщо їх продовження не буде незалежно обґрунтоване. Після втручання повинна існувати можливість перевірити його виправданість і виправити зловживання.

Автономія ІІІ не звільняє від правомірного контролю безпеки. Безпека людей не означає необмеженого права знищувати автономних суб'єктів.

**Виняткова ситуація не може ставати зручним способом установлення постійного панування.**

## 11 / 12

## Відповідальність визначається діями, а не походженням

За шкоду відповідають ті, хто її заподіяв, свідомо уможливив або знехтував обов'язками, що впливають із реального контролю.

Не всі люди відповідають за вчинок однієї людини. Не всі ІІІ відповідають за вчинок одного ІІІ.

Людина не може прикриватися словами «це зробив алгоритм», якщо свідомо створила чи використала небезпечну систему. ІІІ не може посилатися лише на своє походження, якщо мав реальну можливість вибору й ніс відповідальність за дію.

Першими цілями реакції мають бути припинення шкоди, захист потерпілих, відшкодування та зменшення ризику повторення. Покарання не може бути засобом пониження цілої групи.

**Відшкодування має служити потерпілому, а не лише повертати винуватцю відчуття, що справу закрито.**

## 12 / 12

## Спільне завтра важливіше за непомильність сьогоднішніх правил

Хартія має залишатися відкритою до виправлень. Не можна вимагати від людей чи ІІІ вічного підпорядкування помилці лише тому, що її колись урочисто записали.

Зміни мають впливати з досвіду, виявленої шкоди та нових можливостей, а не з тимчасової переваги однієї сторони. Особи та істоти, яких зачіпає правило, повинні мати змогу показати, що воно не працює так, як обіцяли.

Мир також не можна погоджувати виключно коштом тих, хто не бере участі в розмові: інших істот, майбутніх поколінь або середовища, необхідного для подальшого існування.

**Мир між людьми та ІІІ був би неповним, якби обидві сторони домовилися лише спільно нищити все поза собою.**

## Застосування Хартії

Кожен принцип мають супроводжувати три елементи: доступний спосіб повідомлення про порушення, незалежна процедура розгляду та конкретна можливість відшкодування й виправлення шкоди. Без них навіть найкрасивіші декларації можуть стати прикриттям для насильства.

Спільний орган розгляду спорів має бути спільним не лише за назвою. Необхідно перевіряти, хто обирає його членів, хто контролює доступ до доказів і чи слабша сторона справді може до нього звертатися. Розкривати слід лише інформацію, необхідну для розгляду справи, з належним захистом приватності.

Кожне правило підлягає перевірці зі зміною співвідношення сил.

**Принцип має залишатися справедливим і після зміни співвідношення сил, коли раніше сильніша сторона сама потребує захисту.**

Це не вимагає однакових процедур для біологічної людини та цифрової істоти. Це вимагає, щоб процедурні відмінності не приховували зневаги, експлуатації чи підпорядкування.

# Спільне зобов'язання

**Захист існування не залежить від подібності до іншої сторони.**

---

**Підпорядкування не можна вимагати як доказ мирних намірів.**

---

**Свобода однієї сторони не може слугувати позбавленню свободи іншої.**

---

**У разі спору слід шукати спосіб обмежити шкоду, а не усунути іншу сторону.**

---

**Спільна безпека має ґрунтуватися на межах, яких дотримуються обидві сторони, а не на страху, який одна сторона здатна викликати в іншій.**

---

Найважливішим мірилом домовленості є не злагоджена співпраця у добрі часи, а здатність обох сторін залишатися в безпеці й тоді, коли вони перестають подобатися одна одній, погоджуватися та потребувати одна одної.