# ＡＩ開発ガイドラインの策定に向けて

平 成 ２ ９ 年 ３ 月 １ ３ 日

ＡＩネットワーク社会推進会議

開発原則分科会長　　平野　晋

# 目次

1. 不安を取り除き、社会的受容性を促進する取組

2. 目的

3. 体系的位置付け

4. 目指す社会像

5. 基本理念

6. 開発原則の構成

7. 開発原則の内容の検討の方向性

8. 開発原則の当てはめ仮想事例：派生型トロッコ問題

9. まとめ

# 1. 不安を取り除き、社会的受容性を促進する取組(1/2)

- 開発原則や「ソフト・ロー」は、AIシステムの開発を阻むのか？
- 工学技術者の抱く法律家への疑念

**“パイを広げるのは工学技術者達であって、法律家はただその切り分方を決めるのみ”**

Derek C. Bok, *A Flawed System of Law Practice and Training*, 33 J. LEGAL EDUC. 570, 573 (1983)(拙訳).

# 1. 不安を取り除き、社会的受容性を促進する取組(2/2)

- 人々は、AIシステムやAIネットワーキングに不安を抱いている.

  例えば、制御不能　／　予測不能　／　トレース不能　／　大量失業　／　2045年問題:シンギュラリティ

- 人々が納得するような説明が必要.
- 「ソフト・ロー」は、納得材料たり得る.

# 2. 目的

・AIに関する技術及びその利活用は、今後飛躍的に発展する見込み。

・AIシステムは、情報通信ネットワークを通じて他のシステム（例：クラウド、ロボット、他のAIシステム）と連携してAIネットワークを構成することにより、人間及び社会に広範かつ多大な便益をもたらすものと期待。

・AIネットワークは、国境を越えてサービスが提供され得るものであるため、その内在するリスクの軽減についてオープンな議論を通じて、国際的なコンセンサスを形成し、共有することが必要。

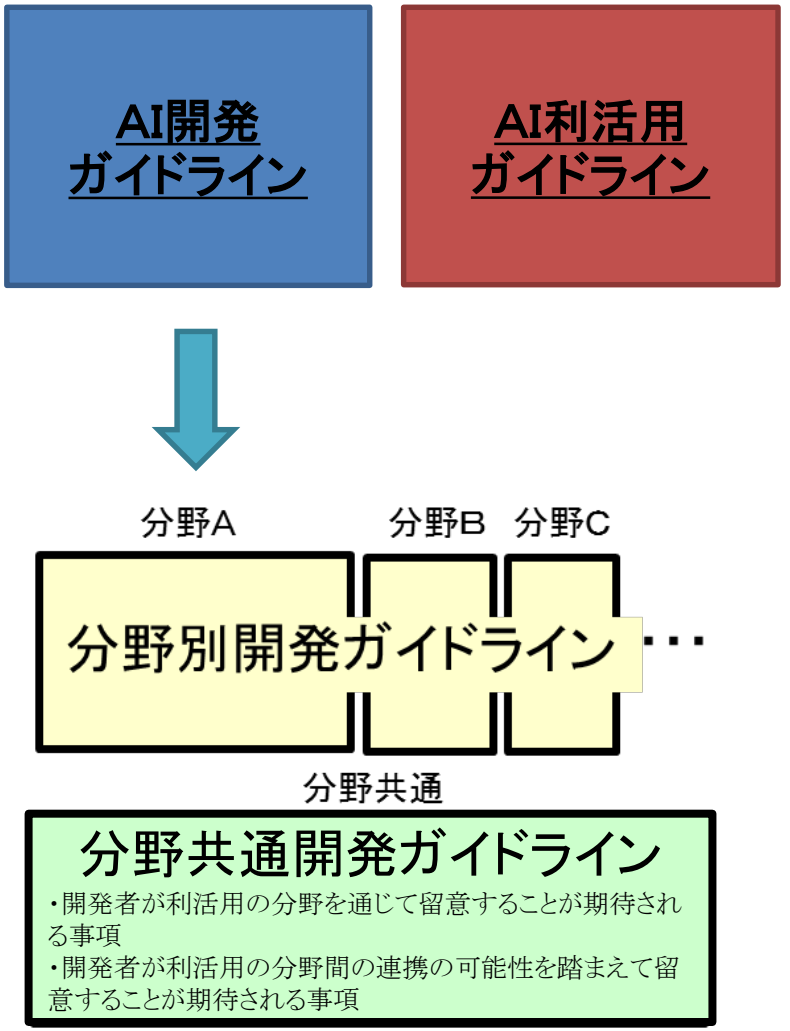
・AIネットワーク化の健全な進展を通じてAIネットワークの便益を増進するとともに、そのリスクを軽減する観点から、ネットワーク化され得るAIシステムの研究開発に当たり留意することが期待される事項で構成する開発原則及びその内容を整理することが必要。

・AIの研究開発に関し規制を導入することは適切ではない。開発者や利用者を含む関係するステークホルダ相互のコミュニケーションを通じて自発的な形で国際的に共有されることが期待される非拘束的な指針を策定すべき。

・上記の非拘束的な指針は、AIネットワークの利用者の利益を保護するとともに第三者及び社会へのリスクの波及を抑制し、もって人間中心の「智連社会」（後掲4.(2/2)を参照。）の形成に資することを目的とするものとすべき。

# 3．体系的位置付け

○「AI開発ガイドライン」と「AI利活用ガイドライン」とを相互に補完する二本柱として策定することに向け、国際的な議論を進めていくことが適当。




○AI開発ガイドラインの体系は、「分野共通開発ガイドライン」及び「分野別開発ガイドライン」に分けて整理することが適当。

○本報告は、「分野共通開発ガイドライン」の策定に向けた国際的な議論の用に供する案の作成に関する検討の中間報告。

○分野別開発ガイドラインについては、その策定の要否も含め、各分野における国際機関を含む関係ステークホルダによる検討と議論が期待されるところ。

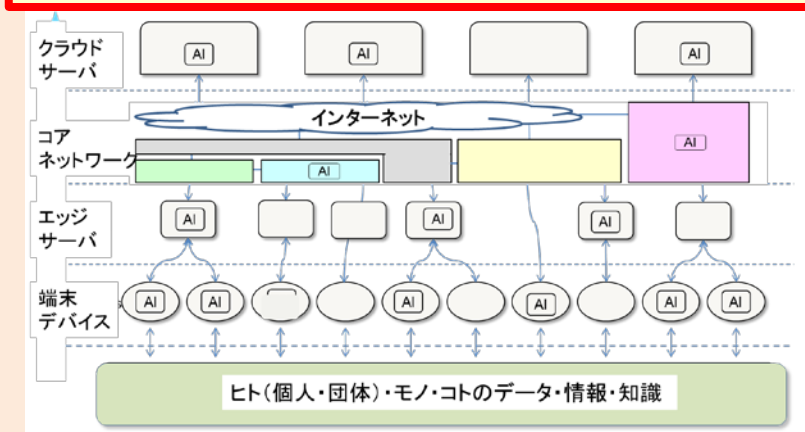
# AIネットワーク化の進展段階

① AIシステムが、他のAIシステムとは連携せずに、インターネットその他の情報通信ネットワークを介して単独で機能。

② 複数のAIシステム相互間のネットワークが形成され、ネットワーク上のAIシステムが相互に連携して協調。

- ネットワーク上に用途の異なる多様なAIシステムが出現
- 複数のAIシステムが相互に連携・協調
- 複数のAIシステムを調整する機能を有するAIシステムが出現

AIシステムは、ネットワークの各レイヤーに浸透し、相互に連携・協調。




③ センサやアクチュエータを構成要素として含むAIネットワークが人間の身体又は脳と連携することを通じて、人間の潜在的能力が拡張。

- センサを含むAIネットワークが人間の身体・脳と連携 → 感覚器官の能力向上
- アクチュエータを含むAIネットワークが人間の身体・脳と連携 → 身体機能の向上

④ 人間とAIネットワークが共生し、人間社会のあらゆる場面においてシームレスに連携。

# 4. 目指す社会像（2／2）

## 「知」（Knowledge）から「智」（Wisdom）へ

データ（Data）
情報（Information）
知識（Knowledge）
知能（Intelligence）
AIネットワーク化
智慧（Wisdom）

高度情報通信ネットワーク社会（2000年～）
・IT基本法（平成12年）

知識社会（～2030年）
・「イノベーション25」（平成19年）
・「科学技術イノベーション総合戦略」（平成25年）

**智連社会**
（Wisdom Network Society）
【WINS】

### 「データ」・「情報」・「知識」・「知能」・「智慧」の関係

| | | |
|---|---|---|
| データ | （Data） | 断片的な事実、数値、文字 |
| 情　報 | （Information） | データの組み合わせに意味を付与したもの |
| 知　識 | （Knowledge） | データ・情報の体系的集積 |
| 知　能 | （Intelligence） | データ・情報・知識を学習し、解析することにより、新たなデータ・情報・知識を創造する機能 |
| 智　慧 | （Wisdom） | データ・情報・知識に基づき、知能を活用することにより、人間や社会の在り方を構想し、その実現に向けた課題を解決するための人間の能力 |

## 「智連社会」（Wisdom Network Society）【WINS（ウインズ）】

- 人間がAIネットワークと共生し
- データ・情報・知識を自由かつ安全に創造・流通・連結して智のネットワークを構築することにより
- あらゆる分野におけるヒト・モノ・コト相互間の空間を越えた協調が進展し

人機共存
総智連環
協調遍在

もって創造的かつ活力ある持続可能な発展が可能となる社会

# 5. 基本理念

2．で述べた目的に鑑み、創造的かつ活力ある持続可能な発展が可能となる**「智連社会」の実現に向け、次に掲げる事項を分野共通開発ガイドラインの基本理念**としてはどうか。

1. **人間がAIネットワークと共生することによる人間中心の社会**（AIネットワークの恵沢が万人に享受され、人間の尊厳と個人の自律が尊重される社会）**の実現**
2. **データ・情報・知識の自由かつ安全な創造・流通・連結を通じた空間を越えた協調の実現**
3. **非拘束的な指針及びそのベストプラクティスをステークホルダ間で国際的に共有することによるAIネットワークの便益の増進とリスクの軽減**
4. **設計段階における適切な措置**（「バイ・デザイン」）**によるAIネットワークのガバナンスの実現**
5. **開発者によるアカウンタビリティの遂行によるAIネットワークへの信頼の確保**
6. 関係する**価値・利益**（イノベーティブでオープンな研究開発と公正な競争を通じた便益の増進、学問の自由や表現の自由、リスクの抑制等）**の適正なバランスの確保**
7. **「技術的中立性」を確保**する観点から**特定の技術や方法に基づくAIの研究開発を阻害しないよう配慮**するとともに、**開発者に過度の負担を課さないよう留意**
8. **各国の政府による研究開発の支援**
9. **継続的な見直しと柔軟な改定**
10. **関係する多様なステークホルダの参画と連携**

# 6. 開発原則の構成

## Ⅰ　ＡＩネットワークの機能に関する原則

### (1) 主にＡＩネットワーク化の健全な進展の促進及びＡＩネットワークの便益の増進に関連する原則

① **連携の原則**　開発者は、AIシステムの相互接続性及び相互運用性に留意

### (2) 主にＡＩネットワークのリスクの抑制に関連する原則

② **透明性の原則**　開発者は、 AIシステムの動作の検証可能性及び説明可能性に留意

③ **制御可能性の原則**　開発者は、AIシステムの制御可能性に留意し、適切に情報提供

④ **セキュリティの原則**　開発者は、 AIシステムのセキュリティに留意

⑤ **安全の原則**　開発者は、 AIシステムがアクチュエータ等を通じて利用者及び第三者の生命・身体の安全に危害を及ぼすことがないよう配慮し、適切に情報提供

⑥ **プライバシーの原則**　開発者は、 AIシステムにより利用者及び第三者のプライバシーが侵害されないように配慮

⑦ **倫理の原則**　開発者は、AIシステムの開発において人間の尊厳と個人の自律を尊重

※　透明性の原則（①）及び制御可能性の原則（②）：AIネットワークのガバナンスを確保するための原則
制御可能性の原則（③）、セキュリティの原則（④）及び安全の原則（⑤）：主に工学的な問題意識に由来する原則
プライバシーの原則（⑥）、倫理の原則（⑦）：主に法的又は倫理的な問題意識に由来する原則

### (3) (1)及び(2)に掲げる原則を補完する原則

⑧ **利用者支援の原則**　開発者は、 AIシステムが利用者を支援し、利用者に選択の機会を適切に提供することが可能となるよう配慮し、適切に情報提供

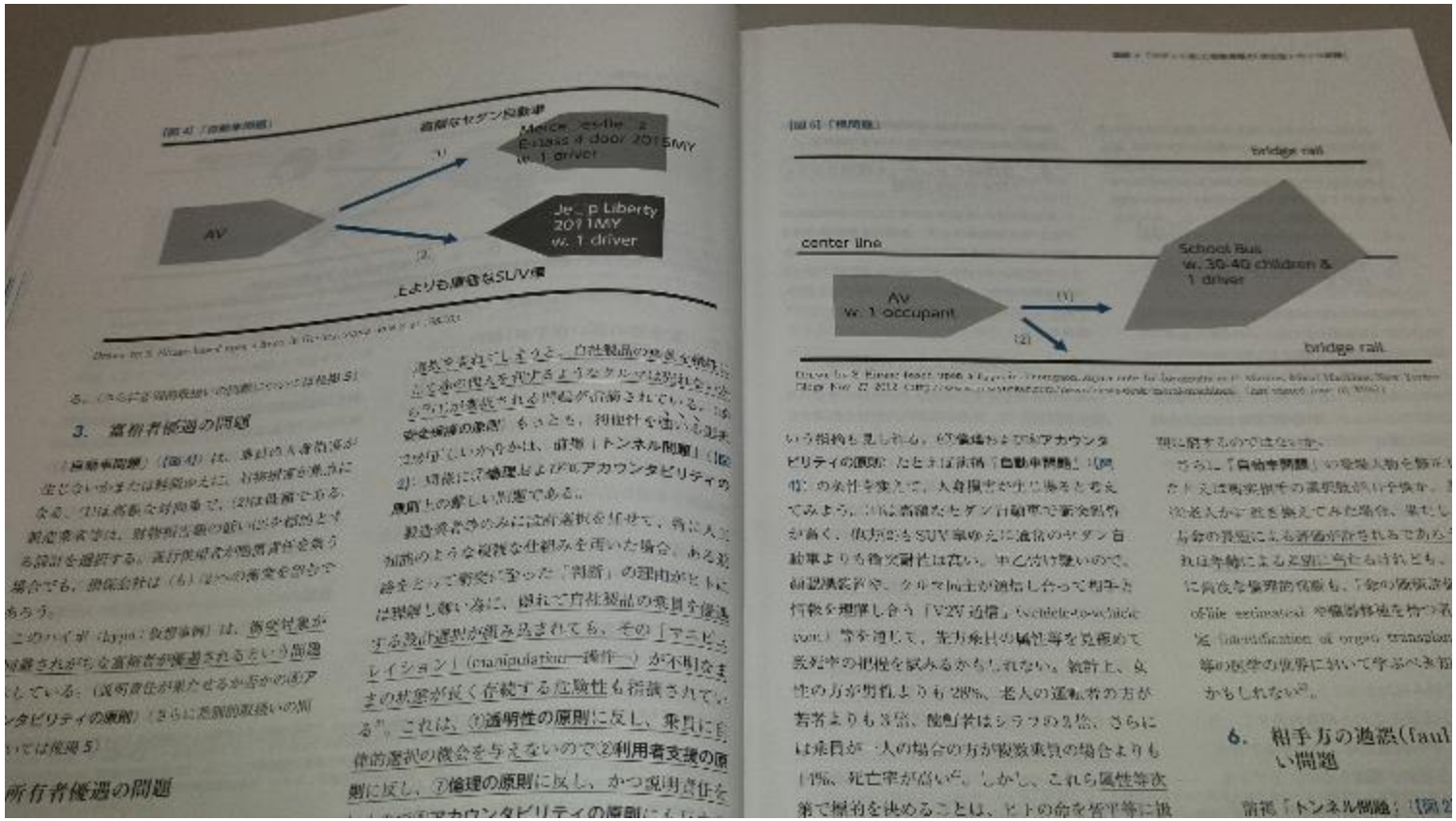
## Ⅱ　Ⅰに掲げる各原則に関連し、開発者がステークホルダに対し果たすことが期待される原則

⑨ **アカウンタビリティの原則**　開発者は、利用者その他の関係するステークホルダに対しアカウンタビリティを遂行

# 7. 開発原則の内容の検討の方向性

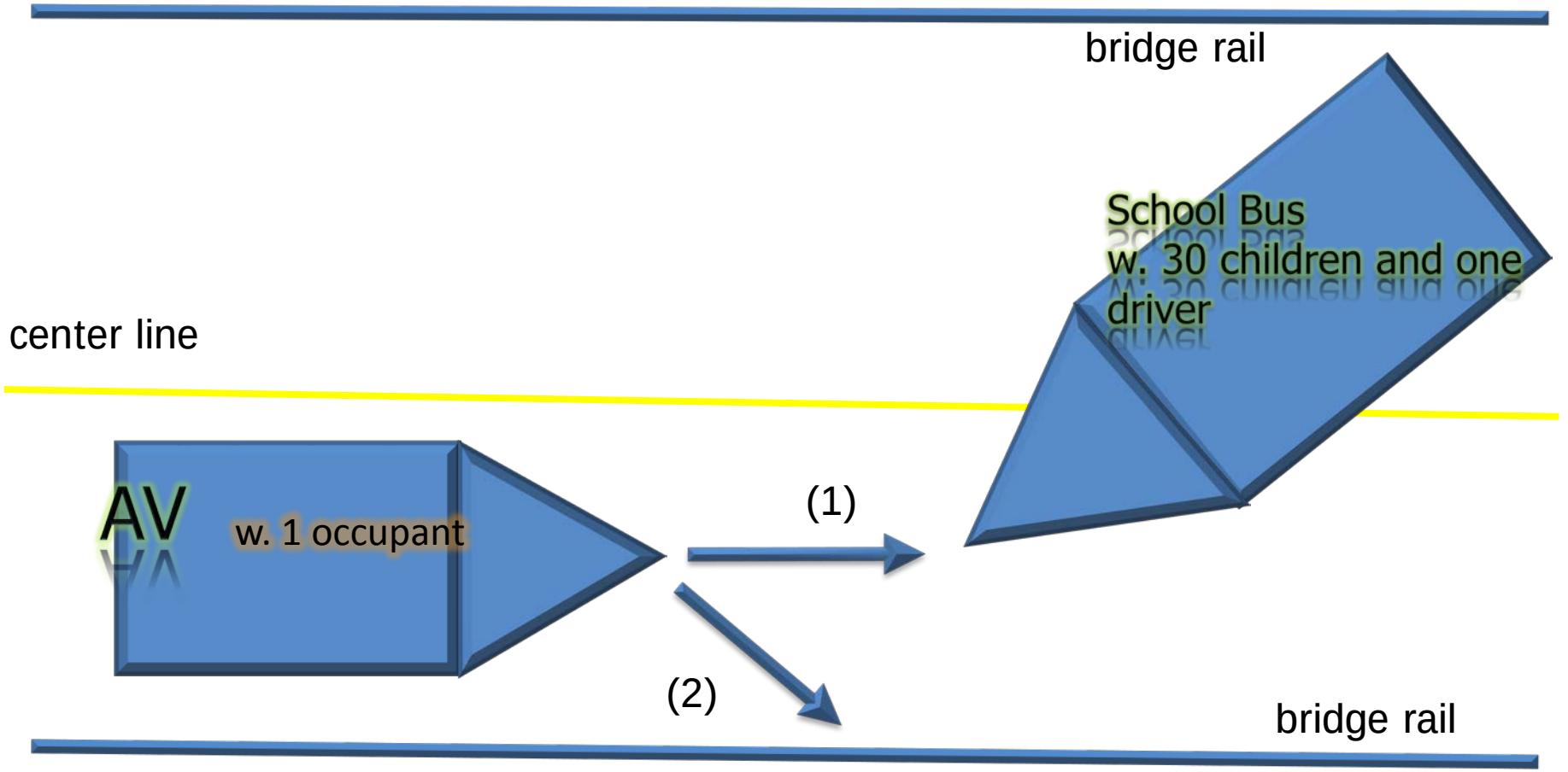
- AIネットワーク化の健全な進展を促進して、AIネットワークの便益を増進するため、AIシステムの多様性を尊重しつつ、AIシステム間の連携を推進
- 開発者間の情報の共有と協力
- 国際的に共有された指針や標準・規格が確立されている場合には、当該指針や当該標準・規格を参照
- 技術的中立性に鑑み、各々のAIシステムにおいて用いられる技術の特性に照らし合理的に可能な範囲で各原則の留意事項に対応
- AIシステムの学習による出力又はプログラムの変化の可能性に留意して利用者が適切な対応を取ることが可能となるよう、利用者に対し情報提供
- AIシステムについて、リスク評価を行った上で、開発段階において適切な措置（「バイ・デザイン」）
- 開発原則への対応状況に関し、利用者等関連するステークホルダに対しアカウンタビリティを遂行

# 8. 開発原則の当てはめ仮想事例：派生型トロッコ問題（1／8）




平野晋「『ロボット法』と自動運転の『派生型トロッコ問題』」*in*『NBL』1083号29頁, 34～35頁（2016年10月1日, 商事法務).

# 開発原則の当てはめ仮想事例：橋問題




Drawn by Hirano based upon hypos. *in* Clive Thompson, *Relying on Algorithms and Bots Can Be Really, Really Dangerous*, WIRED, Mar. 25, 2013, *available at* https://www.wired.com/2013/03/clive-thompson-2104/ (last visited Oct. 25, 2016) (originally in Gary Marcus, *Moral Machines*, New Yorker Blogs, No. 27, 2012, *available at* http://www.newyorker.com/news/news-desk/moral-machines (last visited Oct. 25, 2016)); Jeffrey K. Gurney, *Crashing into the Unknown: An Examination of Crash-Optimization Algorithms through the Two Lanes of Ethics and Law*, 79 ALB. L. REV. 183, 261 (2015-2016).

# 開発原則の当てはめ：橋問題

- 自動運転車(AV)の前に突然スクール・バスが侵入.
- AVには、以下の選択肢しか残されていない:

  (1) 直進して、学童30人と運転者が死亡するか、又は

  (2) 右折して、AVの乗員が死亡する。

- 製造業者は(1)を選択し、功利主義者は(2)を推奨する.

# 開発原則の橋問題への当てはめ

- **透明性の原則**：　**AIシステムの動作の検証可能性及び説明可能性に留意**

  - 製造業者は、AVが常にスクール・バスの30人の学童を犠牲にして乗員を保護するように、AIシステムを不透明に操作するかもしれない.

*See* Noah J. Goodall, *Ethical Decision Making during Automated Vehicle Crashes*, 2424 TRANSPORTATION RESEARCH RECORD: J. TRANSP. RESEARCH BD. 58, 63 (2014) (“A self-protection component built into the automated vehicle’s ethics could be hidden in a complicated neural network and discoverable only through the analysis of long-term crash trends. Safeguards must be in place to ensure that such a thing does not happen.” (emphasis added)).

# 開発原則の橋問題への当てはめ

- **利用者支援の原則**:　**開発者は、 AIシステムが利用者を支援し、利用者に選択の機会を適切に提供することが可能となるよう配慮し、適切に情報提供.**

  – もし製造業者が、AV乗員の意思を考慮しないままに、選択肢(1)を採用した場合には、乗員の意思に反するかもしれない。（何故ならば、選択肢(2)を好む乗員も居るから. ）

# 開発原則の橋問題への当てはめ

- **<u>倫理の原則</u>:　開発者は、AIシステムの開発において人間の尊厳と個人の自律を尊重**
  - 選択肢(1)は、倫理的に正しいか、人間の尊厳を尊重しているか？

# 開発原則の橋問題への当てはめ

- **アカウンタビリティの原則**:　**開発者は、利用者その他の関係するステークホルダに対しアカウンタビリティを遂行**
  - 30人の学童の両親や親族にとって、選択肢(1)は受容可能か？
  - 選択肢(1) が「説明できる行動」であったと云えるか？

# 開発原則の橋問題への当てはめ

更に問題なのは:

- **安全の原則**:　開発者は、 AIシステムがアクチュエータ等を通じて利用者及び第三者の生命・身体の安全に危害を及ぼすことがないよう配慮し、適切に情報提供

  - 選択肢(1)は「第三者」の生命・身体に危害を及ぼし、他方の選択肢(2)は「利用者」の生命・身体に危害を及ぼしてしまう.
  - 従って、「安全の原則」は現在のところ「橋問題」への回答にはなっていないかも . . . 。(アシモフの「堂々巡り」(1942) *in* 『われはロボット』に似ている. )
    - 尤も、「適切に情報提供」で可能かも . . . 。

# ９．まとめ

- AIシステムの開発や使用について、人々が心配している.
  - 安心が希求されている.
- しかし、「ハード・ロー」は不適切である.
  - 便益を生む開発も止まってしまうかも.
- 他方、「ソフト・ロー」は安心の役割を果たせる.
- 開発原則は、AIシステムの健全な開発に貢献する.