

## SFと現実の境界を越えるAIロボット

### 栗原 聡

慶應義塾大学理工学部 慶應義塾大学共生知能創発社会研究センター オムロン サイニックエックス株式会社





#### 理想と現実(アカデミアの立場から)



安全安心について考える

理想

人間の自助による解決 人間同士の対話による解決

現実

自助は無理 ←テクノロジーの発展が起因 現状は思考停止状態←黒船が来ないと変われない? 周りはテクノロジーを活用してくる

→結局は活用せざるを得なくなる





#### なぜAIを活用するのか?



現場においてのAI活用 センシング性能 ≫ 人 →活用すべき (認識の速度と精度においてAIは人を凌ぐ)

判断においてのAI活用←生成AIを活用することになる

分析能力・推論能力・判断能力において人を凌駕する

- ※緻密な反実仮想シミュレーション能力を有する
- →<u>基本的に活用すべき(周りは活用するはず)</u>
- ※現状だとハルシネーション対策・インジェクション対策などで懸念あり



#### 2019年段階での分類

#### 想定通りに展開されてしまっている.



タイプA 人がトリガー を引き一連の 動作において も人が直接 的に操作す る道具として の兵器・武器

#### 拳銃など



タイプB 人がトリガーを ひくが、途中 の過程に多く が自動化され た兵器 (直接操作感 は薄れるがトリ ガーは人が引

トマホーク, 自動追尾型 ミサイルなど (エアコンなどの自動 家電もココ)

#### 現在開発されている段階はココ

タイプC プログラムに書かれた手 順に従いトリガーをひく 動作を含め自動的に 目的を実行する自律 型兵器(自動兵器) ※プログラム時点で想 定されない事態には対 応しない (出来ない)

ルンバ型 <機械が人を殺傷する図式 に見える>

※司令する側としては想定 通りに動作する.

※司令レベルで想定されな い事態に対応することもない。

タイプD 複数のタイプC型の自 律型兵器が連携する マルチ自律兵器

個々のAIレベルでの動作は規定 されていても、複数AIの協調に おいて想定外の事態が発生する 可能性がある(フラッシュ・クラッシュ). ※兵器としては致命的となり得る.

タイプC'

プログラムに書かれた手 順に従い状況に応じて 複数の目的の中から, 実行する目的を自動 設定するを自律型兵 器(自動兵器)

※プログラム時点で想 定されない事態には対 応しない(出来ない)

汎用性ルンパ型 ルンバ型が状況に応じて 巧妙に動作を変化させつつ 作戦を実行する.

あたかも意思を持って自律 的に動作するように見える. ※基本はタイプC型と同様

複数のタイプC'型の自

律型兵器が連携する

想定外の事態が起きる

マルチ自律兵器

可能性が増す.

タイプD'

タイプF 複数のタイプE型の自律 型兵器が連携する マルチ自律兵器

もはや人の手に負えないレベル

非現実的

危険地帯, 劣悪環境等での展開として有望だが, 要注意.

タイプE 与えれた目的の達成

のため, 自らが自律 的に動作を決定する 自律型兵器

ターミネータ型

<機械が意図を持って人を殺傷する 図式に見える>

※司令する側としては楽. 100%制御可能性が保障されない 兵器の使用はありえないか?

※一般社会が想像する自律型兵器はココ

※人と共生するAIもこのレベル

参考【タイプG】 与えれたメタ目的の達 成のため, 自らが自 律的に目的を生成し 実行する高レベル自 律型AI



Super Intelligence \

想定通りに展開され てしまっている.

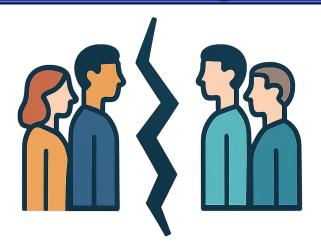






### インターネット(SNS)・AI がもたらしつつあるもの





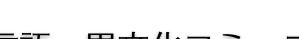
デマ,フィルターバブル,フェイク,二極化, 炎上,ネットいじめ,陰謀論,,・・・・・・

サイバー空間に止まらず、日常生活自体も巻き込む

- <膨大な量・多様な質>が人の認知能力を大きく越える
  - →SNSの登場が人の思考力を低下・脊髄反射的反応を助長
  - →安易な生成AI多用→思考力の低下・多様性低下・適応力低下・

→<br/>
不寛容增長・利己的・<br/>
社会性低下

現状のテクノロジーの発展がもたらすもの



異言語・異文化コミュニケーションは大切だ! とは言うが・・・ →現実は、同質化・均質化→異なるモノの排除に向かう





#### テクノロジーの進化:テクノロジーが及ぼす人間社会の臨界











引用画像出典:「政党要件という両刃の剣…なぜ「れいわ新選組」はテレビに映らないのか」、毎日新聞、2019年7月19日掲載

https://mainichi.jp/articles/20190719/k0 0/00m/010/177000c



インターネット→煩悩増幅器フィルターバブルアテンションエコノミーナラティブ指向個人の心理→群集心理化※民主主義の崩壊

地球環境問題 国際協調〔外交·紛争〕

エケシム

政治 少子高齢化 2040年問題 介護問題 生産性低下

生産性低ト イノベーション力低下・・・

いやがらせ・ストーカー対策専門窓口「群 集心理により引き起こされた『複数人による ストーカーや嫌がらせ』」 https://fam-iyagarase.com/crowdpsychology-stalker/







### 人間の自助では対応不可能な状況→真剣にまずい.









テクノロジーの加速的な発展がもたらす人間社会の臨界



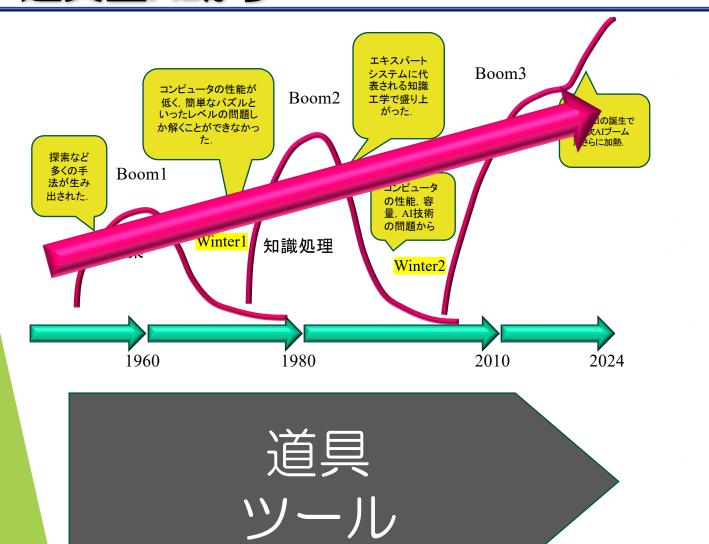






### 道具型AIから・・・・・・

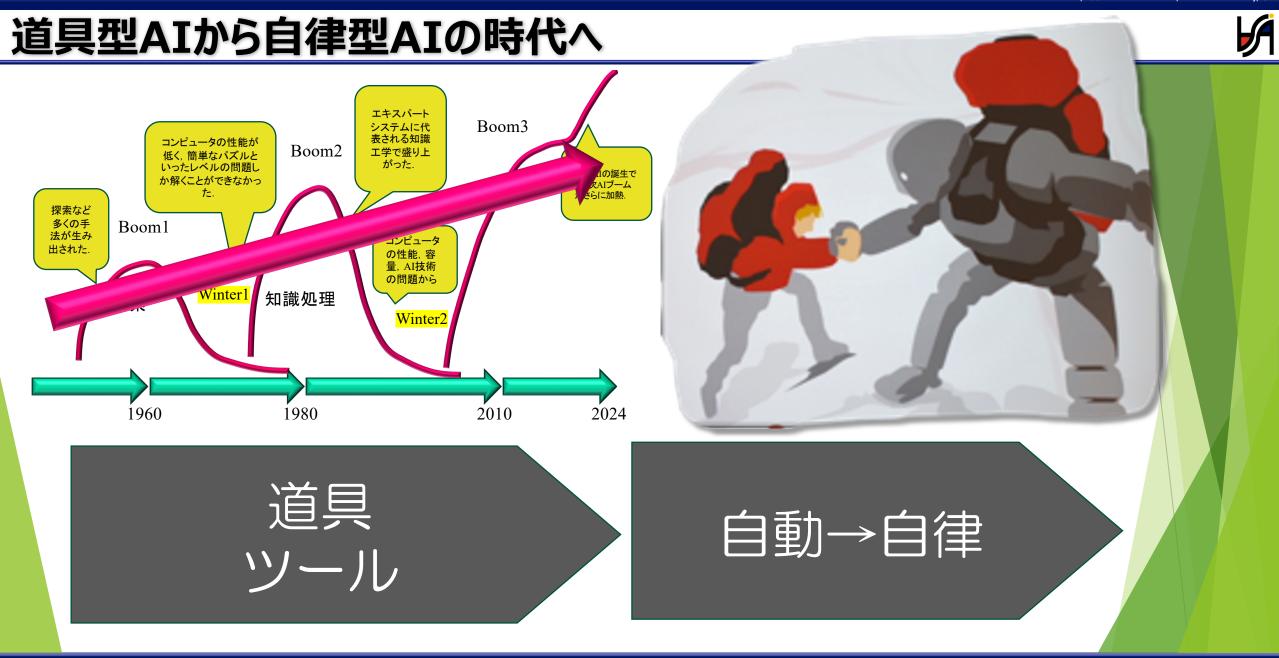
















#### いよいよAIはSFの世界へ



- AIは道具から自律にシフトしていく
  - →最終的にトリガーはAIに任せることになる.
- 相手もAIに判断させるとしたら・・・・
  - →瞬時の判断が必要
  - →人の方が判断を誤り判断に時間がかかる.
- ◆厳格な自律型AIの運用ができるのか?
- ◆そもそもAIへの目的の設定の仕方は?





#### 結果がすべてであるのであれば・・・・・



厳格運用を徹底すれば→適応力が低くくなる

- →現場での適応能力が高い方が有利
- →AIにすべてを委任するようになるしかない.

AIシステムの到達点が完全自律分散型

- →高い適応能力を持つ.
- →人は抽象度の高い命令のみとなる
  - ※AIの人への説明は必須



#### 技術的な懸念



- →群れでの想定外の創発への対応
- ◆自律型AIにおける目的設定変更不可対応
- →学習能力
  - →勝つために学習することの行く先







# AIの認識力と 判断力

牛成AI活用

道具型AIから 自律型AIへ

単独AIから 群AIへ







#### 自分(アカデミア)への問い



DUAL Useとしての安全保障のための 研究開発とは?

DUAL Useを意識した研究開発とは?

単に意識するだけでは変化は起きない

- →DUAL Useを具体的に意識した研究設定
  - ※関係者との対話が必要



