AI時代の技術倫理

361°アートワークス Vトーク配信
アバター倫理学のための10のガイドトーク
技術論・シンギュラリティ編

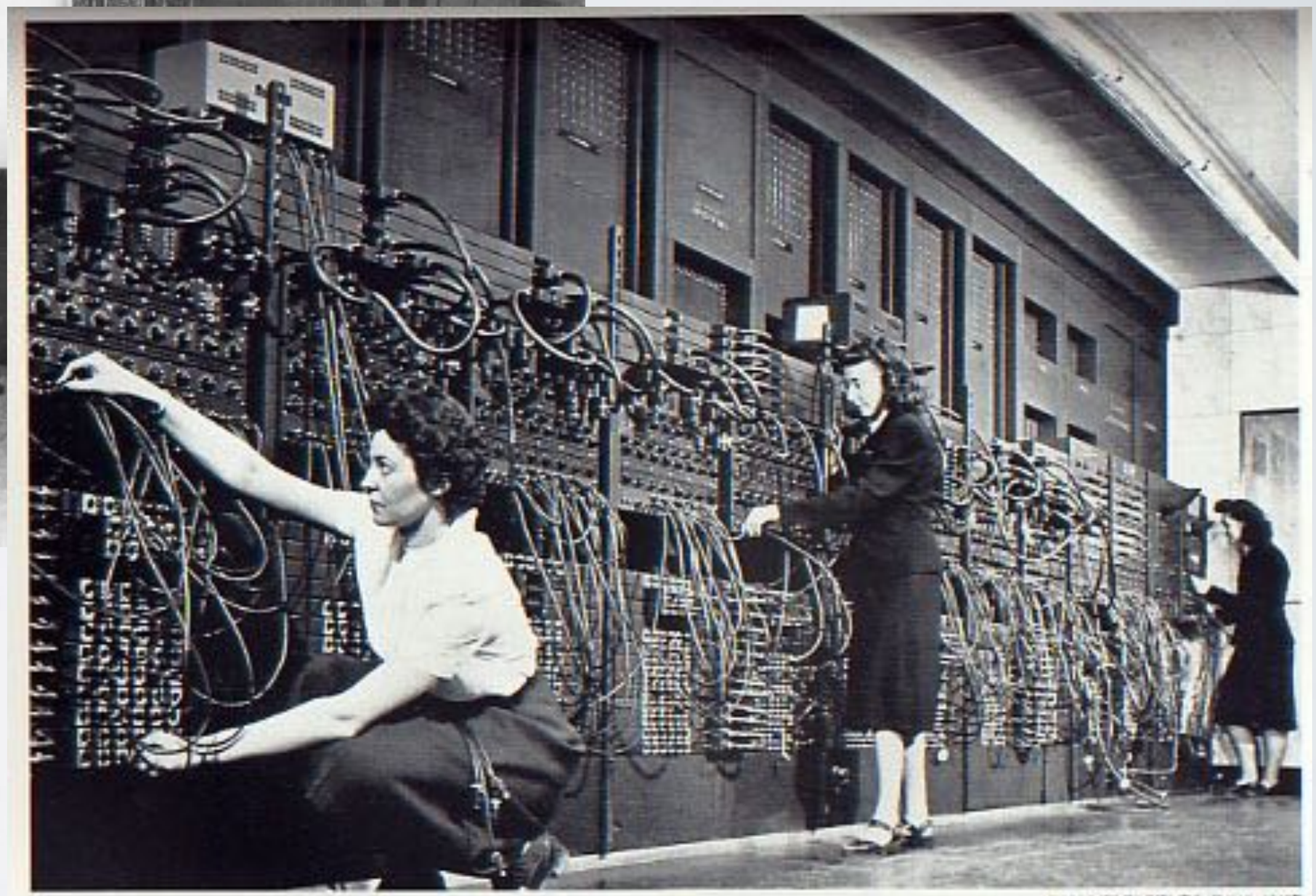
HTTPS://WWW.YOUTUBE.COM/CHANNEL/UCMIZFJNGQGGO4ZFZJV1UP0Q/LIVE

2019.9.21 放送

浸透いうな先生
(インターネット崩壊論者)

1946 ENIAC (世界最初の汎用電子計算機?)

ここから今日まで74年



ENIACのプラグを接続し直す、アメリカ、1946年

1963 Sutherland / Sketchpad

ここから今日まで56年

1964 IBM System/360

メインフレームの時代

最大1.5MIPS 主記憶最大8MB

1969 ARPANET

インターネットのはじまり

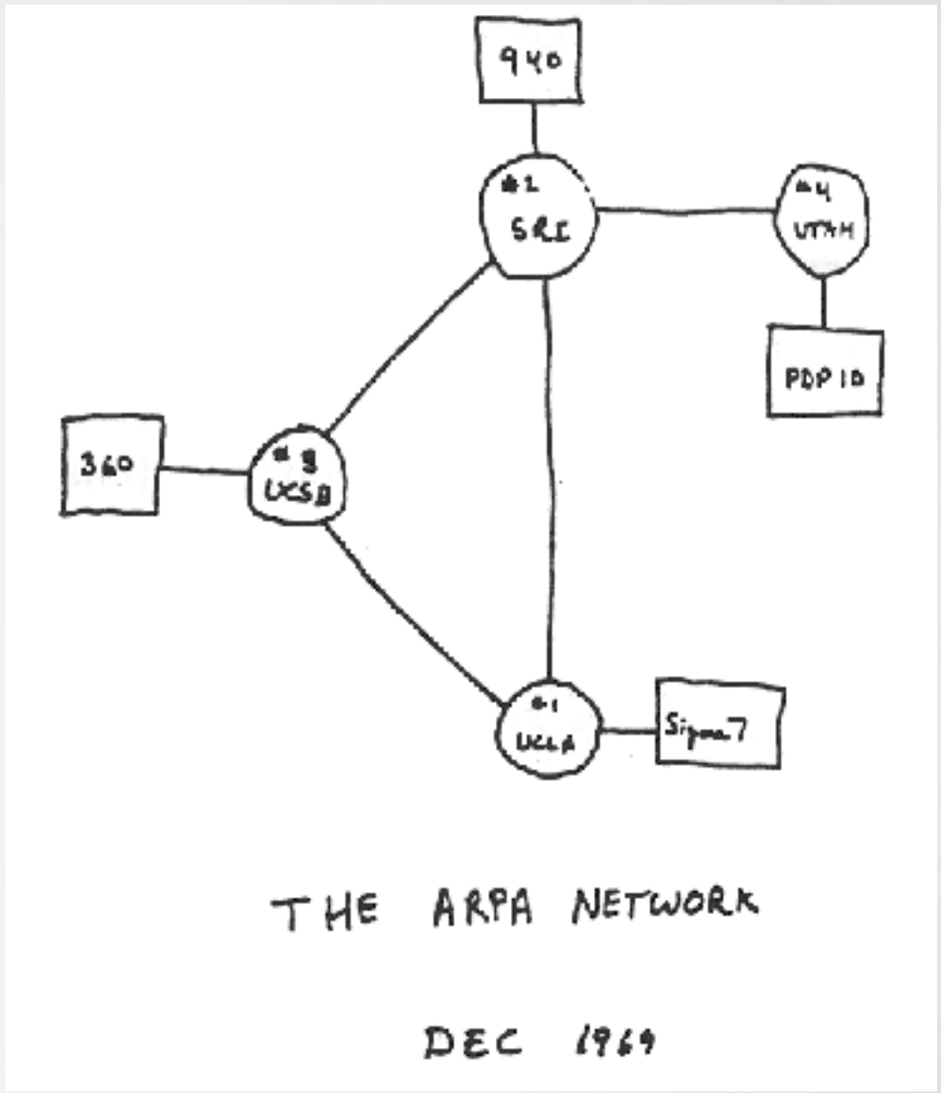
凄いアイデアだな。
１００万ドル使っていい。がんばれ。

第2代ARPA局長チャールズ・ハーツフェルド

規格の異なるコンピュータを結ぶ
ネットワークを構築したい。(1966)

IPTO部長職 ロバート・テイラー

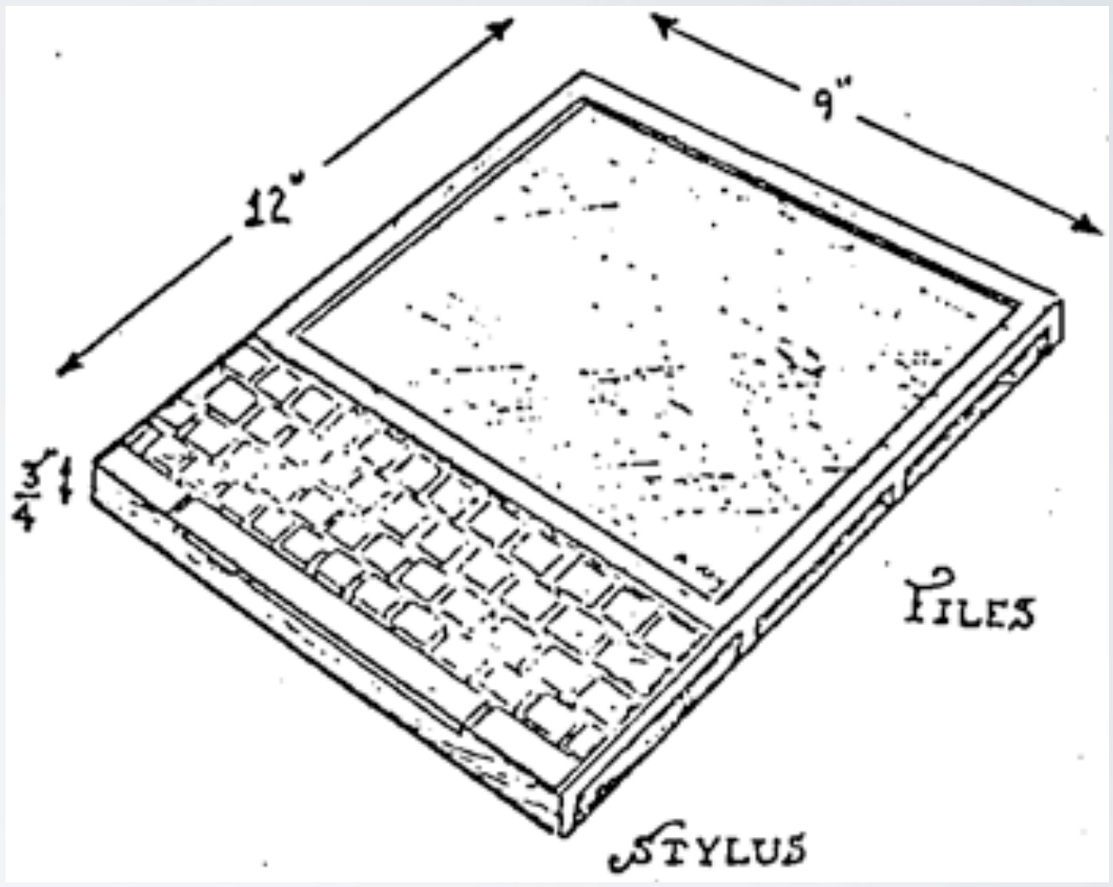
ここから今日まで50年



1972 Dynabook / Alan Kay

“あらゆる年齢の「子供たち」のための
のパーソナルコンピュータ” 構想

ここから今日まで47年



「コンピュータが子供をプログラムするべきか、子供がコンピュータをプログラムするべきか？」

ー S. パパート

1977 Apple][

パソコンの登場

ここから今日まで42年

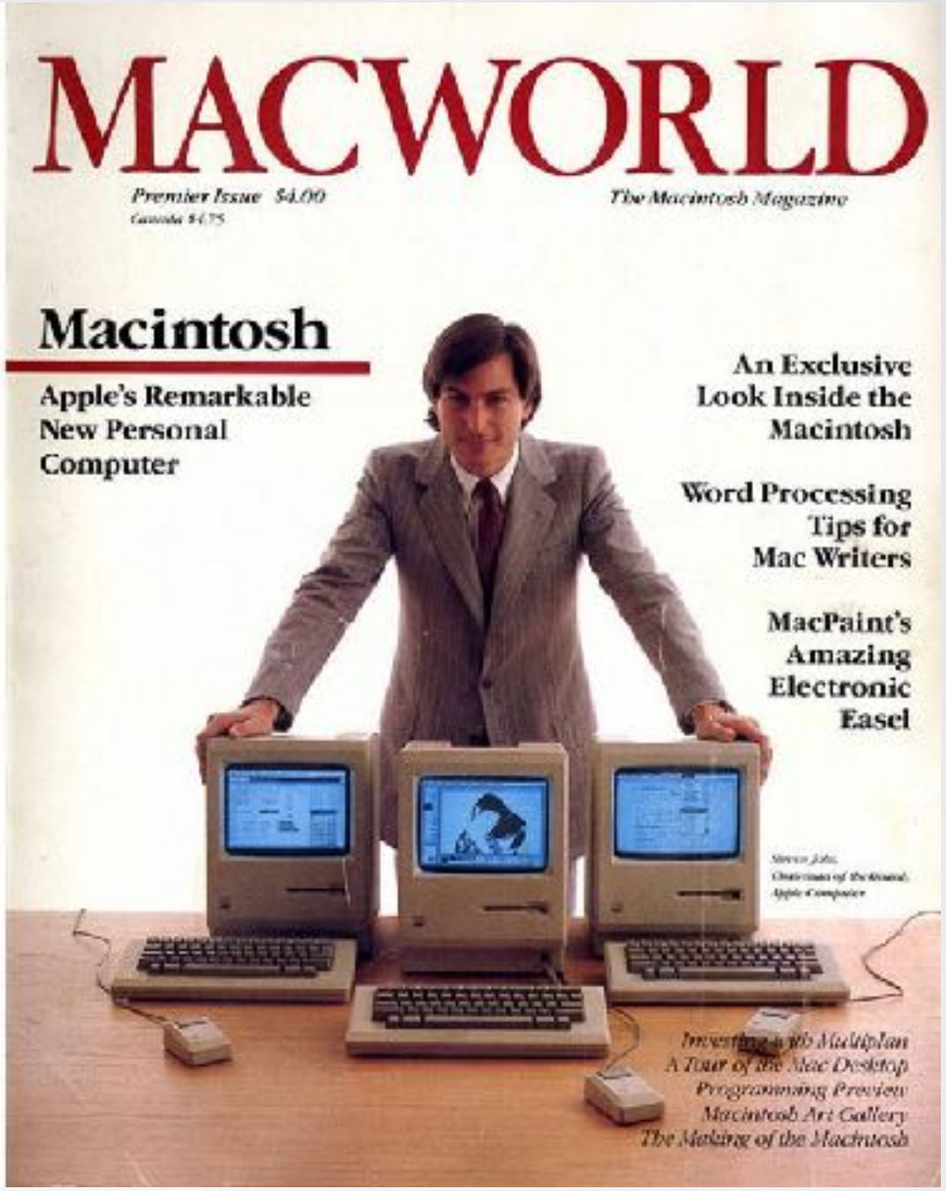
1979 ALTO / Smalltalk

ここから
今日まで
40年

(GUI ワークステーションの登場)

1984 Macintosh

ここから今日まで35年



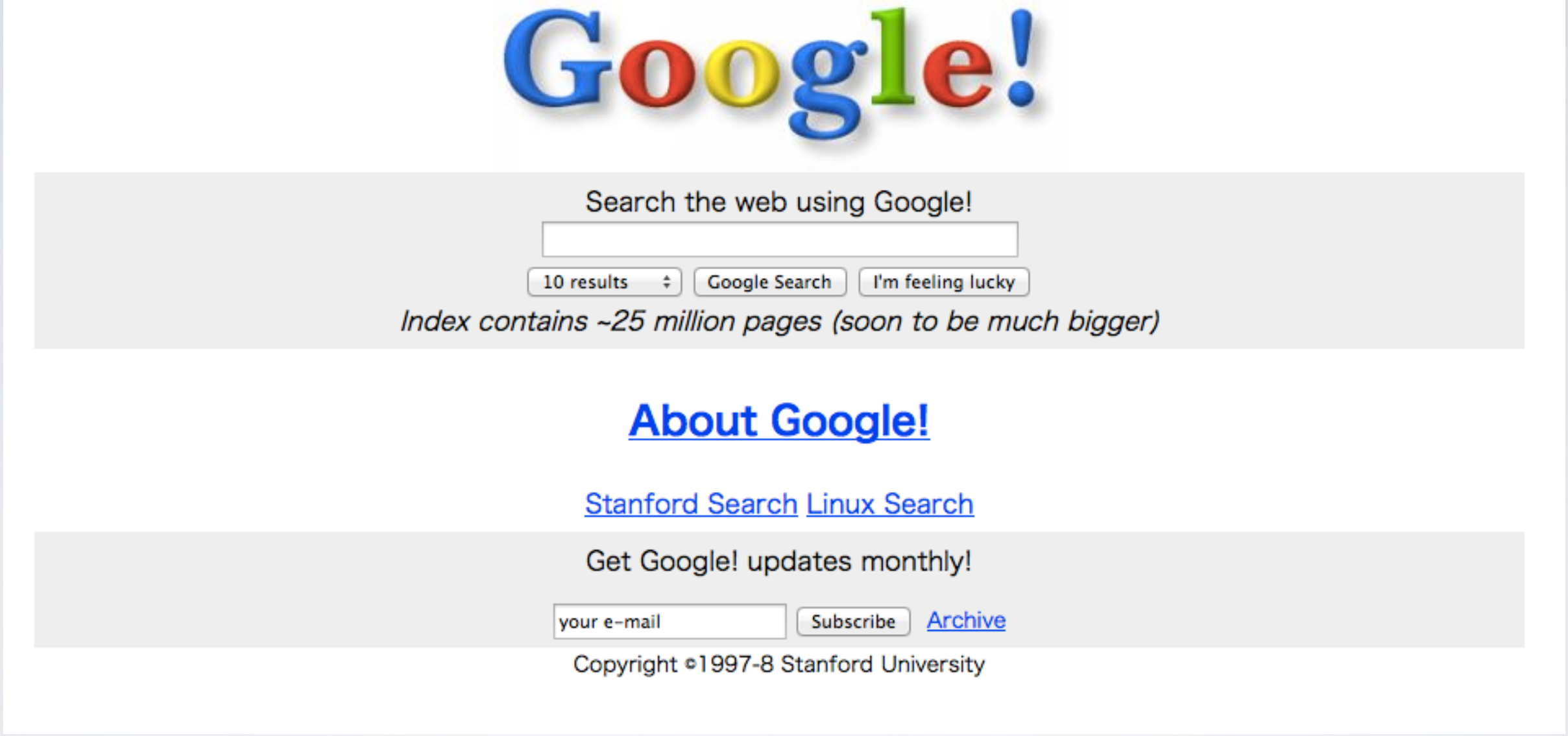
(GUI を備えたパーソナルコンピュータの登場)

1995 Windows95

(市民がインターネットに繋がる時代へ)

1998 Google

ここから今日まで21年

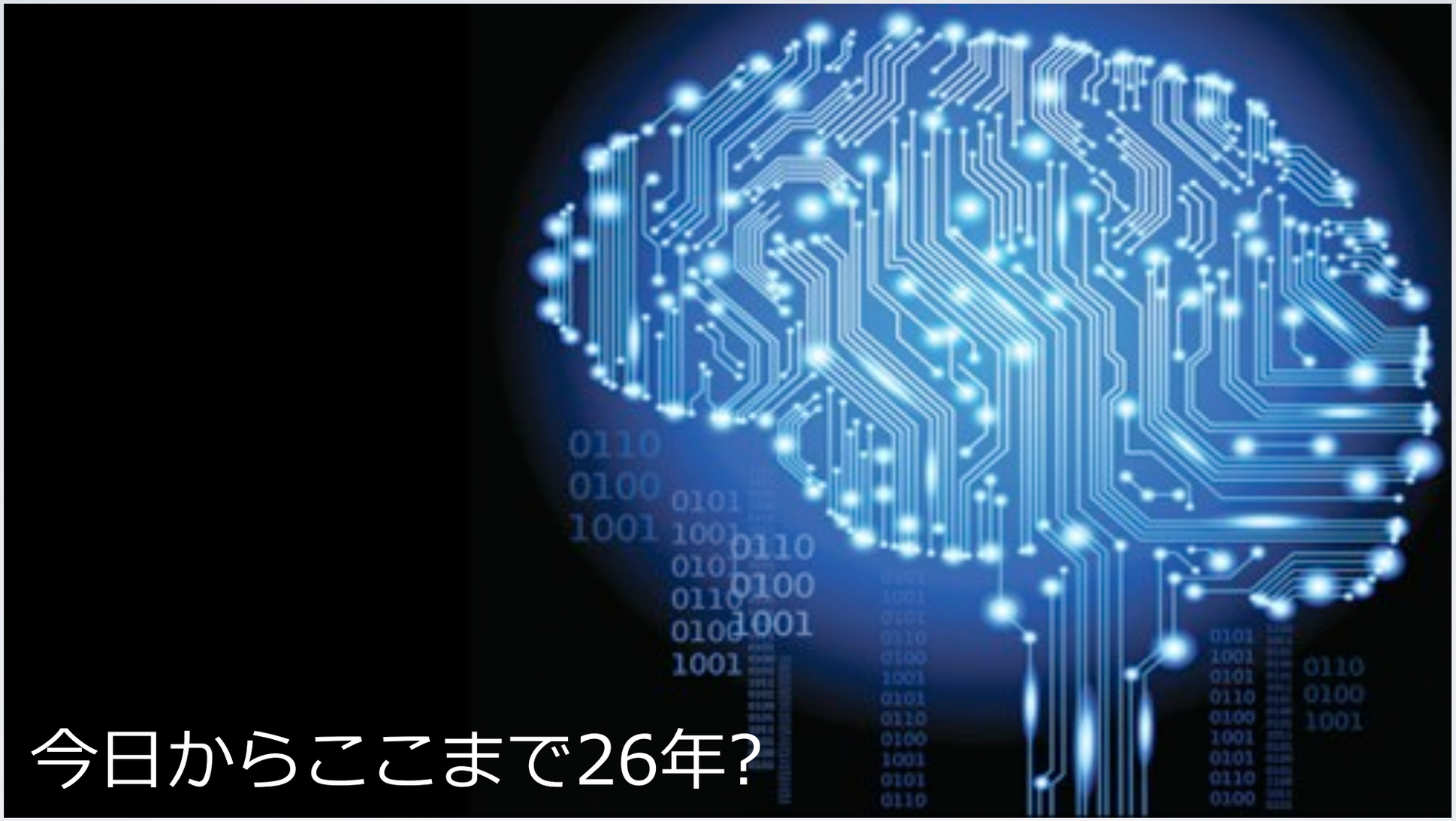


(検索の時代へ)

2007 iPhone / 2010 iPad

(コンピュータは創造から消費の道具へ?)

2045? シンギュラリティ (技術的特異点)



レイ・カーツワイルの予想

『ポスト・ヒューマン誕生 - コンピュータが人類の知性を超えるとき』

1,000ドルのコンピュータが全人類の知性を追い越す

AI が自身を進化させ始める

シンギュラリティについて考えてみよう

- 人工知能を破壊できるか? (『動物マークⅢの魂』)
- 人工知能に恋をしたら? (『Ex Machina』)
- 完全な記憶を持った存在とか (Life loggger)
- 発言の矛盾をいちいち指摘される世界とか
- 情報処理のみならず科学が一気に進む世界

(だからシンギュラリティ(技術的特異点)という)

(不老不死とか / ブレインアップロードとか)

シンギュラリティについて考えてみよう

- 雇用はどうなる? プログラマは不要になる?
- 不労社会が訪れたとき経済システムはどうなる?
- 誰が統治する社会になるのか
  - 国や独裁者から AI あるいは AI の独占企業へ?
  - AI が奴隷なら奴隷を所有するのは誰か?
  - あるいは我々が AI の奴隷になるのか? (まさか? もうなりかけてるでしょ?)
- AI を制御できるか? 閉じ込められるか?

ジェイムズバラット　人工知能人類最悪にして最後の発明

ユドカウスキーの懸念 フレンドリーな AI は可能か

なぜ未来は我々を必要としないのか？ nano machine (ビルジョイ)

http://miyata3.com/Great_Essay/Why_the_future_doesnt_need_us.html

われわれは遺伝子機械として組み立てられ、ミーム機械として教化されてきた。しかしわれわれには、これら創造者にはむかう力がある。この地上で、唯一われわれだけが、利己的な自己複製子たちの専制支配に反逆できるのである。 (Dawkins 1976 利己的遺伝子)

シンギュラリティについて考えてみよう

“ In the end
they will lay their freedom at our feet
and say to us,
'Make us your slaves, but feed us.' “

Dostoevsky

民主的AIが必要だ (WBAI代表 山川宏)

「私は、ＡＩの成果は特定の組織などが独占するのではなく、共有財産にすべきものだと考えています。多くの人々に富を分配するかたちにしたいのです。」

木村 草太,佐藤 優,山川 宏. AI時代の憲法論 人工知能に人権はあるか (Japanese Edition)

人工知能学会倫理指針

3（他者のプライバシーの尊重）
4（公正性）人工知能学会会員は、人工知能の開発と利用において常に公正さを持ち、人工知能が人間社会において不公平や格差をもたらす可能性があることを認識し、開発にあたって差別を行わないよう留意する。人工知能学会会員は人類が公平、平等に人工知能を利用できるように努める。
5（安全性）
9（人工知能への倫理遵守の要請）

総務省 AI 研究の開発に関する 8 原則

「透明性」「利用者支援」「制御可能性」「セキュリティ確保」「安全性」「プライバシー保護」「倫理の尊重」「アカウンタビリティの原則」

http://www.soumu.go.jp/joho_kokusai/g7ict/main_content/ai.pdf

シンギュラリティは来るのか

- コンピュータが創造、発見をし始める?

  → それはすでに始まっている

- コンピュータは感情を持つか? 感情とは?

- 動物は感情を持たない?
- 人は簡単に感情移入する

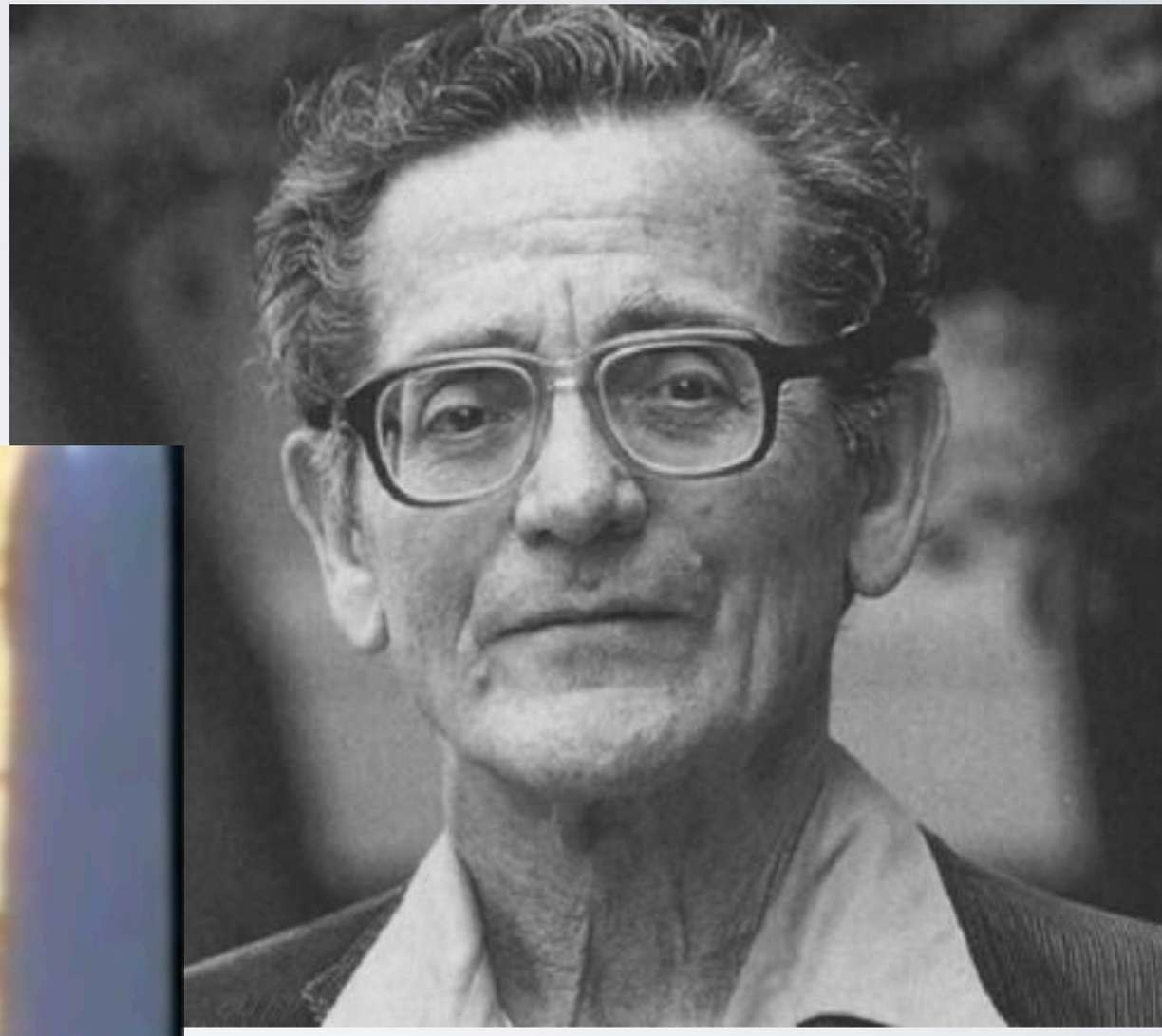
  動物マークIII、たまごっち、アイボ

シンギュラリティは来るのか

- そもそも知能とは? 意識の問題とは区別
- 貴方のような意識をコンピュータは持てないのか?
- そもそも貴方には意識はあるのか?
  - 動物は意識を持っていないのか?
- ある鳥が鴨のように見え、鴨のように泳ぎ、鴨のように鳴くならば、それはたぶん鴨なのだ。
- 問題は自意識、自我 (私とは?)
- 高い知能、自我の芽生えには身体性が必要 → VR

自由意志は存在するか

次スライドなし



- M.ガザニガ『社会的脳』

- B.リベット
  『マインドタイム』

自由意志は存在するか

神経細胞は1000億、視覚だけでも1000万ビット/秒

→意識が生ずる前に膨大な量の情報の処理が必要

→意識はリアルタイム処理できない

→意識は0.5秒程度の遅れをごまかされている

→結果を説明させられるのが意識

→思考?

　言語化した思考はとても遅い / ひらめきとは

自由意志は存在するか

- 前野隆司『脳はなぜ「心」を作ったのか「私」の謎を解く受動意識仮説』

　神秘的にみえる意識は幻想、人間の脳も無数のエージェントの情報処理で動いている

→ 人工知能は可能

(ただし量子脳理論がハードルか)

技術はどうあるべきか

- のび太のテロ
- E.F.シューマッハー
  『スモール イズ ビューティフル』
- I. イリイチ
  『コンヴィヴィアリティのための道具』
- 崩壊するインターノット
- 木村草太, 山川宏 他
  『AI時代の憲法論 人工知能に人権はあるか』

技術はどうあるべきか

- のび太の悪行
  - ドラえもんの道具を勝手に使った 347回
  - 道具でしずかちゃんにいたずら　542回
  - 地球を滅亡させそうになった　207回

のび太のテロ

- どこにでもいるのび太やドラえもん
  - 京大カンニング事件
  - バイトテロ
  - あおり運転
  - 個人情報ばら撒きサーバ
  - 放置される数百万台の欠陥ルータ
  - 暴露されるプライバシー (一億総盗撮者)

→ リテラシーだけの問題か

→身の丈にあわない技術の問題ではないか

E.F.シューマッハー

- スモール イズ ビューティフル (1973)
  - 今日の多くの技術は「民衆の利害」によって普及がはかられているのではなく、「科学者、大企業、政府の利害」によって進路が決められている
  - 科学と技術がやがて、もうすぐすべての人を豊かにするので、余暇と財産を一体どうあつかうかという問題だけが残り、他の問題が消滅してしまうだろうというのは幻想

E.F.シューマッハー

- スモール イズ ビューティフル (1973)
  - 大規模組織は往々にしてきわめて行儀が悪く、反道徳的で、愚劣でかつ人間性にもとる行動にはしるが、それはそうは組織内部の人間の性格によるのではなく、ただ組織が巨大さの重みを引きづっているからなのである。
  - 組織の中の人間は行動の自由が無いので道徳的な挫折感をいだく

E.F.シューマッハー

- スモール イズ ビューティフル (1973)
  - 機械化には二種類ある
    - 第一は、人間の技能と能力を高める機械化
    - 第二は、人間の仕事を機械という奴隷に引渡し、人間をその奴隷への奉仕者にしてしまう機械化

E.F.シューマッハー

- スモール イズ ビューティフル (1973)

  - 現代経済学：多く消費する人が消費の少ない人より豊か 適正規模の生産努力で消費を極大化

  - 仏教経済学：理想は最小限の消費で最大限の幸福を得る 適正規模の消費で人間としての満足を極大化

I.イリイチ

- Ivan Illich

  1926.9.4～2002.12.2

- Tools for Conviviality 1973

- メキシコ
  クエルナバーカ
  CIDOC(国際文化資料センター)

I.イリイチ

- 自然な規模と限界
- 限界を超えたとき
    - 新たな奴隷制
    - 校舎化、病棟化、獄舎化
- 科学技術が管理する人々にではなく、相互に結びついた個人に仕えるような社会が "コンヴィヴィアル"

I.イリイチ

- 豊める国々の囚人はしばしば、彼らの家族よりも多くの品やサービスが利用できるが、品物がどのように作られるかということに発言権を持たない し、その品物をどうするかということも決められない。彼らの刑罰は、私のいわゆるコンヴィヴィアリティを剥奪されていることに存する。彼らは単なる消費者 の地位に降格されているのだ。

I.イリイチ

- コンヴィヴィアルな道具とは、それを用いる各人に、おのれの想像力の結果として環境をゆたかなものにする最大の機会を与える道具のことである。

I.イリイチ

- 災厄の避けがたい暴威を、自立共生的な再構築へと革命的に転回させるいとなみに、大多数の人々を協同させることができるものは、弱きものとしての言葉あるのみなのだ。

Rise of the Stupid Network

単にビットを運ぼう、愚直に！
(JUST DELIVER THE BITS, STUPID)

エンドユーザ装置は自由に柔軟にふるまって良い．なぜなら，Stupid Network においてはデータが殿様であり，bit は本質的に自由であり，そして，データが単一のデータ速度やデータタイプであるといった仮定は存在しないからである。

August 1997

David Isenberg / 訳 佐藤正和

VIEWS OF THE FUTURE

D.D.Clark said

We reject: kings, presidents and voting.
We believe in: rough consensus and running code

1992 24th Internet Engineering Task Force

World of Ends

Doc Searls and David Weinberger

インターネットはものではない。合意である

インターネットの3つの利点

a. 誰もそれを所有していない
b. 誰もがそれを使える
c. 誰でもそれを改善できる

崩壊するインターノット

- 基盤技術は進化していない
  - TCP/IP (RFC 793/791)：1981年(38年前)
  - SMTP (RFC 821,822)　：1982年(37年前)
  - DNS (RFC 1034,1035)：1987年(32年前)
  - BGP4 (RFC 1654)　　：1994年(25年前)
  - IPv6 (RFC 1883)　　：1995年(24年前)

崩壊するインターノット

- 自律の喪失から規制強化へ
- 分散運用が立ち行かなくなり集中管理へ
- 協調の喪失から分断へ
- 生み出す力の喪失 (J.ジットレイン『インターネットが死ぬ日』)
- 自由からの逃亡 (安心のファシズム)
- 利便性と安全性の追求からプライバシーの喪失へ
  (偏見や差別や抑圧が無くなるならよいが)
- インターネット(＝自律分散協調)の崩壊
- インターノットも根幹が同じままで動き続けられるか?

崩壊するインターノット

- 砂上の楼閣である現実を直視し崩壊に備えよう
  - 崩壊シナリオA: 機能不全
  - 崩壊シナリオB: 管理ネットワーク/管理社会 (自律分散協調の敗北)
- どういうネットワーク、どういう社会にしたいのか? (現状は哲学不在)
- 管理される社会か、
  自律・分散(選択の自由)・協調の社会か、、、

AI がどうなるかにも左右される、自律分散かAI に管理される社会か