情報分野での研究倫理

平木　敬

東京大学理学部情報科学科

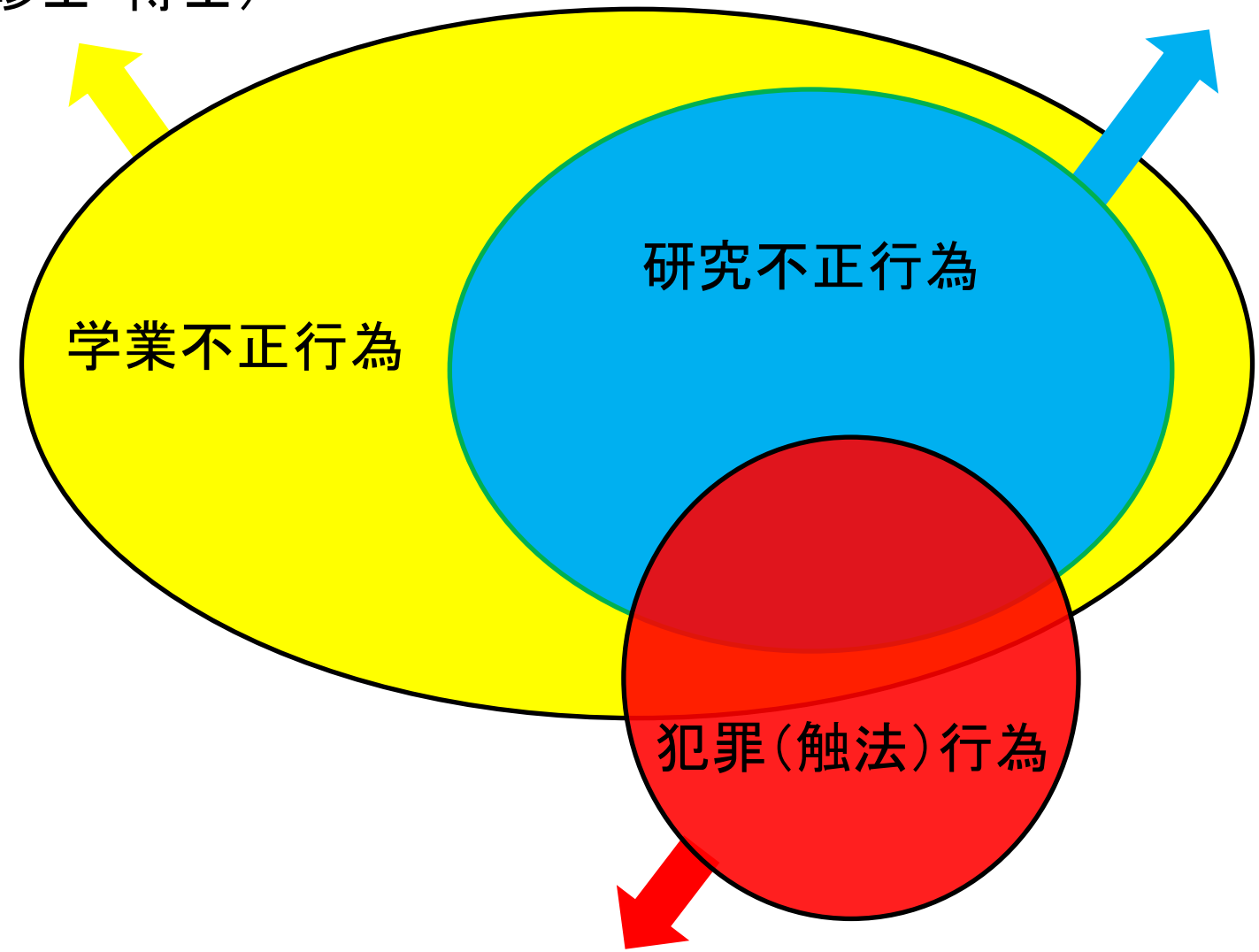
研究・学習・生活における注意点

- 研究倫理
  - 誤った研究をしない
  - 不適切な論文を作成しない
- 学業倫理
  - 不適切な手段で単位を取得
  - 研究倫理を外れた実験・研究
  - 不適切な学位論文
- 社会的倫理
  - 触法行為をしない
  - 他人の迷惑となることをしない

様々な不正行為

学位剥奪・退学
（学士・修士・博士）

懲戒・研究費停止
（論文・学会発表）



罰金・刑務所

研究不正行為

- 誤った方法による研究実施
- 誤った方法による論文執筆

- FFP　最も重大な研究不正
  - 捏造　（Fabrication)
  - 改ざん（Falsification
  - 盗用（剽窃）　（Plagiarism）

故意によるものではないことが根拠をもって明らかにされたものは不正行為には当たらない[1]

FFPについては、萩谷先生の講義に含まれています

注1）http://www.mext.go.jp/b_menu/shingi/gijyutu/gijyutu12/houkoku/attach/1334660.htm

Research Misconducts1

- **Fabrication**
  - making up data or results and recording or reporting them.
- **Falsification**
  - manipulating research materials, equipment, or processes, or changing or omitting data or results such that the research is not accurately represented in the research record.
- **Plagiarism**
  - the appropriation of another person's ideas, processes, results, or words without giving appropriate credit.

Research misconduct does not include honest error or differences of opinion.

https://www.dhs.gov/xlibrary/assets/foia/mgmt_directive_10500_research_misconduct.pdf

FFP（復習）

- 捏造
  - 行ってない実験を記載
  - 無いデータを作る
- 改竄（かいざん）
  - 都合の悪いデータを変更する
  - 条件を変えて書く
- 盗用（剽窃）
  - 他の文章、図を自分のものとしてコピペする
  - 引用（引用の明記、変更なし、適切な分量）は良い

QRP（復習）

- Questionable Research Practice
  - ギフトオーサーシップ
  - 不適切なオーサー順やオーサーからの除外
  - 重複発表や分割発表
  - 先行調査の調査不足、引用不足
  - 自説に有利な実験結果の選択的な発表（改ざんスレスレ）
  - 自説に不利な実験結果の非開示や発表遅れ
  - 予備段階の研究成果の不誠実な発表（とくにメディアに対して）
  - 重要な研究データを、一定期間、保管しないこと
  - 研究記録の不適切な管理
  - 研究試料・研究データの提供拒絶

FFPと比べ、明確な定義がまだ定まっていない

学業不正行為

- 不適切な手段で単位を取得
- 研究倫理を外れた実験・研究
- 不適切な学位論文

- 創作者誤認惹起行為[1]
- 実験結果欺罔行為[1]

} 故意でないミスも含む

- FFP

1）早稲田大学 大学院先進理工学研究科における博士学位論文に関する調査委員会・調査報告書　http://www.waseda.jp/jp/news14/data/140717_committee_report.pdf

創作者誤認惹起行為[1]

- 「創作者誤認惹起行為」とは、真実は自己が創作した文章・図表等又は自己の実験等に基づいて得られた画像・データ等でないにもかかわらず、自己が創作した文章・図表等又は自己の実験等に基づいて得られた画像・データ等であると、読者に誤認させる可能性がある記載を行うことをいう。

- なお、創作者の誤認という結果は、誤認惹起行為に対する認識及び認容（故意）の有無にかかわらず発生する以上、創作者誤認惹起行為該当性の判断について、行為者の故意は要件とはならない。

1）早稲田大学 大学院先進理工学研究科における博士学位論文に関する調査委員会・調査報告書　http://www.waseda.jp/jp/news14/data/140717_committee_report.pdf

創作者誤認惹起行為

- 盗作（剽窃）　　研究不正にも含まれる
- 著作権がない・著作権放棄文献の使用　(古典文書、Wiki系）
- （他人が書いた）報告書などからの転載
- 共著発表論文で、他人が書いた部分の転載
- 代筆
- 業者による翻訳

- 剽窃検知ソフトを悪用した創作者誤認惹起行為の隠ぺい
  - Smith-Watermanアルゴリズムの検出能力は脆弱
- 文章だけでなく、ソフトウェアや説明図の創作者誤認惹起
  - 既存のソフトウェアをつないだだけでソフトウェアを創ったとはいえない

実験結果欺罔行為[1]

- 「実験結果欺罔行為」とは、「真実は自己の実験等で得られた結果やデータ等でないにもかかわらず、読者にそうであると誤信させるために、欺罔の意思をもって、そのように装う行為」をいう。

- 科学論文において、実験結果の記載は、当該論文における作成者の研究成果、作成者の研究者としての能力を示す最も重要な記載であるため、実験結果欺罔行為は科学者としての根本的な価値観・倫理観に反する行為であり、それを行った科学者に対しては、厳しい処分が下されてしかるべきことになる。

1）早稲田大学 大学院先進理工学研究科における博士学位論文に関する調査委員会・調査報告書　http://www.waseda.jp/jp/news14/data/140717_committee_report.pdf

実験結果欺罔行為

- 捏造・改竄　　研究不正にも含まれる
- 他人の実験結果の利用
- （他人が書いた）報告書などからのデータ・写真転載
- 共著発表論文で、他人が実施し書いた実験結果の転載
- 代理実験
- 業者による実験結果

- 学業・学位論文における実験結果欺罔については、まだ定説が固まっていない

社会的不正（犯罪・違法）行為

- 著作権法違反、詐欺など違法行為
- 利益相反
- 許可されていない人間および動物対象実験
- 軍事研究

情報科学分野における研究倫理の特質

- 情報倫理との隣接性
- ソフトウェアの正しさは検証できない
- バグのないソフトウェアはない
- バグと不正の境界が曖昧
- 再現実験は、通常の実験分野より容易
- 「証明」の新規性

実例研究

実例研究の目的

- 不正の身近さを実感してほしい
- 実例のようなことをすると、発覚する

- Disclaimer
  - 直接目撃したケースを含む
  - 卒論、修士論文、博士論文、査読した他組織の論文など
  - 残りは伝聞または完全な創作
  - 特定を避ける為、大きくフェイクが入っている

捏造

捏造1：計算なしでデータを作る

- 研究内容　　スパコンを使って並列計算の性能向上を実現
- 論文では、グラフを使って効果を説明

- 実験をしていない可能性の通報から、Login Logを調査
- 調査結果⇒試にLoginした程度の総Login時間。勿論、実験時間より少ない

即ち、プログラムを作成せず、スパコンを使わないでグラフだけを想像力から作成

捏造

捏造2：評価システムは存在しなかった

- 研究内容　　組み込みシステムのI/O性能の向上
- 論文では、実際に組み込みボードを調達し、グラフを使って効果を説明
- 実験は自宅に組み込みボードを置き、大学では実験しなかった

- データと論文の記述がおかしいことに気付いた
- 研究者の日頃の研究能力と、データを取得するまでの複雑さがマッチしなかった
- 調査結果⇒実験環境の実在性の証拠を求めたが、出せなかった

即ち、組み込みボードによる評価は行われなかった。データはパソコンでの粗いシミュレータで作成

捏造

捏造3：殆ど動かないソフトを評価した

- 研究内容　　共有データベースと協調する協調クライアント群の効果
- 論文では、システムを完成させ、グラフを使って効果を説明
- 作成したソフトは複雑性が高く、動かなかった

- データの数がすくなかった、データが全部似過ぎていた
- 作成したソフトの動作を実演させたところ、データが取れるほど動かなかった
- 調査結果⇒バグを避け、たまたま動くデータセットの結果をアレンジして結果を作った

即ち、バグがとれなかったので、それ自身も不正確な、たまたま動作した時のデータから、他のデータをアレンジ

捏造

改竄

改竄1：性能が悪いデータポイントを改良

- 研究内容　　OS軽量化によるマルチスレッドシステム性能向上
- 論文では、提案方式の全データポイントでの優位性をグラフを使って説明
- 実際には、提案方式の効果は薄かった。都合の悪いデータの書き換え

- 提案内容で考えられないほど、結果が良かった
- 源データを出させたら、論文中のグラフと違っていた
- 調査結果⇒手で、都合の悪いデータを移動。グラフはExcelで書いて、それらしかった

即ち、全部のデータではなく、悪いデータだけを書き換えた

改竄

改竄2：UIの画面例をイラストレータで改善

- 研究内容　　PC上の画面に読みやすい様に変更したWebページを出す
- 論文とプレゼンでは、非常に効果的な表示画面を用いて説明
- 実際には、提案方式の効果は薄かった。都合の悪いデータの書き換え

- 単純なプログラムでは出来ないほど見事な書き換えだった
- デモを要求したら、なぜか簡単なケースしか実施しなかった
- 調査結果⇒実際に作ったソフトの機能は低かったので、イラストレータをつかって画面例を盛った。予定したテストケースしか対応できなかった。

改竄

改竄3：ベンチマークの恣意的選択

- 研究内容　　新しいアーキテクチャ機構の提案と評価
- 論文では、提案方式の優位性をベンチマーク・スイートから選んだもので評価
- 実際には、提案方式の効果が薄いベンチマークを除外して、平均向上率を押し上げた

- 提案内容で考えられないほど、結果が良かった
- 源データを出させたら、結果グラフから除外しているデータがあった
- 調査結果⇒手で、平均値が上がらないベンチマーク結果を削除

即ち、全部のデータではなく、良いデータだけを示した

改竄

剽窃

剽窃1：全体の約半分を剽窃

- 研究内容　　オンラインゲームの高速化
- 論文では、乱数的ツリー探索のアルゴリズム最適化と並列化を扱う
- 実際に研究者がしたのは、クラスタへの実装だけ
- 全体6章のうち、１，２，３章は海外博士論文のほぼストレートコピー
- 図も、ほぼ全てがストレートコピー

- あまりに後半部分と英語の質が違ったので発覚
- 本人は、引用を忘れただけと主張
- 調査結果⇒論文は、3個の論文から大きなブロックでコピペ、図もコピペであることが判った

剽窃

剽窃2：未公表報告書から盗用

- 研究内容　　ネットワーク用アプリケーションソフトの提案と実装
- 研究者が研究グループに入る前からの長いプロジェクト
- 全体6章のうち、１，２章は数年前の研究報告書のほぼストレートコピー
- 当該章にある図も、ほぼ全てがストレートコピー

- 他の部分より滑らかな書きっぷりで発覚
- 審査委員がたまたま報告書の存在を知っていた
- 本人は、引用を忘れただけと主張
- 調査結果⇒論文は、未公表報告書（著者が違う）から大きなブロックでコピペ、図もコピペであることが判った

剽窃

剽窃3:先輩学位論文の再利用

- 研究内容　　人工知能分野
- 研究者が研究グループの一員
- 先輩研究者が博士論文を著述
- イントロダクション、全体のシステム説明、図、写真を先輩の博士論文等からコピペ

- 審査委員が記憶していたことで発覚
- 著作権法違反、厳密な意味での剽窃ではない(許諾があれば)
- 原作者誤認惹起行為にあたる。学位論文には適切でない
- 調査結果⇒他人が著述した文章のコピペは単著の学位論文には適切でない

原作者誤認惹起行為

剽窃4：先輩学位論文の再利用2

- 研究内容　　機械学習応用
- 研究者が研究グループの一員
- 先輩研究者が前年に博士論文を著述
- 研究内容、データ、図が先輩博士論文の一部に酷似

- 審査委員が記憶していたことで発覚
- 文章の剽窃というよりは、アイディア、実験結果の盗用
- 調査結果⇒先輩研究者が当該研究者（修士）のアイディア、実験データを盗用。所属研究室では継続的に行われていた習慣

剽窃・盗用

剽窃5:他人の著作部分の自己剽窃

- 研究内容　　情報科学基礎理論
- 研究者が研究グループの一員
- 先輩研究者が博士論文を著述
- 先輩研究者と共著で論文出版。大部分は博士論文から
- 研究者が書いてない部分を含め、共著論文をコピペ

- 剽窃検知ソフトウェア利用で発覚
- 著作権法違反、厳密な意味での剽窃ではない
- 原作者誤認惹起行為にあたる。学位論文には適切でない
- 調査結果⇒他人が著述した文章のコピペは単著の学位論文には適切でない

原作者誤認惹起行為

剽窃6：定理証明の一部分を先行論文から借用

- 研究内容　　情報科学基礎理論
- 定理証明は、既に既知の部分に研究者が独自の工夫を追加
- 定理証明や、式の展開の一部を先行論文から借用

- 査読者の記憶、参考文献チェックで発覚
- 著作権法違反は微妙（内容の盗用なので）学位論文には適切でない
- 調査結果⇒他人の証明や式の展開の考え方の一部利用なので不適切

剽窃

剽窃7：ほぼ同内容の和文・英文・他国語論文

- 研究内容　　生命シミュレーション
- 1回論文を書くと何回も美味しい
- 防御として、一部データやグラフを入れ替え

- 査読者の記憶、検索によるチェックで発覚
- 論文は学会・出版社にあるので著作権法違反
- 調査結果⇒大部分が重複、言語が違うだけと認定

自己剽窃・重複投稿

実例から：不正を身近な問題と考える

- 不正率は低くない
  - 私の経験からは、約3%の論文に何らかの不正が見つかる
  - 99%の不正は、不正以外の理由で処理される
  - 見つからない不正は多いかは不明
    - ○　ゴキブリ理論、1匹のゴキブリを見ると30匹はいる
    - ×　不正する人はそもそも杜撰なので見つかり易い
  - 剽窃検知ソフトウェアの限界
    - 文章を変える翻案には無力
  - 外部からの投稿論文査読では、より多くのおかしい論文が見つかる
    - 最悪は著者名・タイトル名だけを入れ替えた他人の論文を投稿

なぜ不正をするか

- 不正のメリットは大きい
  - 学位が手に入り就職できる、給与が上がる
  - 業績リストが一つ長くなる
    - 将来の転職に有利
- 不正のデメリットの期待値は残念ながら小さい
  - 最悪ケースで学位取り消し・研究費取消
  - 学業不正 & 研究不正 & 違法行為　ならば無問題
- 最小努力で最大結果を得る最適化の風潮
  - 受験時代の習慣はなかなか抜けない

研究室問題

- 経験から不正発生は均一ではない
  - 不正に甘い研究室はある
  - 不正を暗に要求する研究室もある
  - 不正をしないと出来ない課題を与えられることもある
  - 学生の業績で筆頭著者論文を書く教員はいる
    - 昔は教授が筆頭著者が常識的だった

- 「私の」学位認定基準
  1. 不正が行われていないこと
  2. たとえ達成結果が悪くても、期間中頑張っていたこと
  3. 論文としてキチンと書けている事

- 長い物にまかれてはいけないことを常に確認しよう

おわりに

- 長い目で見れば不正はデメリットばかり
  - 不正し続けて発覚しないことは無理
    - 不正しないと論文が書けない体質になる
    - 多くの事例が語っている
  - 研究者としての実力がつかない
    - 将来の転職が大変

- 不正に気付いたら相談しよう
  - 研究不正の多い研究室は、残念ながらある
  - 相談窓口は沢山ある（なんでも相談室など）
  - 教員や先輩がFFPを示唆したら、駆け込もう

ご意見、誤りなどは hiraki@is.s.u-tokyo.ac.jp までお願いいたします