# AIM Policy: Academic Outsourcing Tell-tales

## 2023 – 2024

**This document outlines common tell-tales of written work that has been 'academically outsourced'. It is always a work in progress…**

## 1 Overview

1.1 As defined in the AIS policy, 'academic outsourcing' is defined as:

- This is unauthorised and/or undeclared human or technological assistance to produce academic work, in whole or part, for academic credit, progression or award, whether or not payment or other favour is involved. It includes, but is not limited to, contract cheating, commissioning, the purchase of pre-written assignments from online repositories ('essay-banks'), inclusion of content provided via self-styled tutorial services, inclusion of content obtained by unauthorised and/or undeclared use of AI-enabled software such as AI text generators ('writers') and paraphrasing tools.

1.2 In terms of tell-tale evidence to look for, academic outsourcing falls into two basic categories:

1.2.1 Text produced by human assignment ghost-writers, possibly working for essay mills, outlined in section 2

1.2.2 Text produced by AI text generators, and other technological tools, outlined in section 3.

## 2 Tell-tales of human ghost-written / essay-mill text.

Indicators that an assignment may have been written by a human ghost-writer, possibly working for an essay mill, include:

2.1 Headline

- document creator (sometimes labelled e.g. 'author') and/or last-modifier (sometimes labelled e.g. 'last saved by') are someone(s) other than the student,

- order / catalogue reference number in filename or on title-page or elsewhere in the document properties / metadata,

  - inspect filenames in the Turnitin assignment inbox for similar filenames that could be arising from students independently commissioning from an essay mill that's penetrated a social-media group,

- essay-mill name or writer ID number appearing as author-name or elsewhere in the document properties / metadata,

- place-holders and/or comments such as 'enter your name here', 'enter assignment title here', 'enter name of placement organisation here' etc.

- an anomalously low similarity score, e.g. a bibliography-excluded score of 0-5%, as aligns with essay-mill 'plagiarism-free' guarantees.

2.2 Style

- Inconsistent writing style or 'voice' in comparison to other work submitted by the student or other available evidence from the student.

- Are the treatment of the subject material and quality of the presentation demonstrably better than, or otherwise demonstrably different to, the student's other assignments?
- Does it look like a supposedly student assignment that's been written by someone with a more advanced or professional command of the subject trying to 'dumb-down' to a learner standard?
- Are the grammar and vocabulary as expected for a student at the level in question?
- Does it look like an assignment that's been written by someone familiar with the subject interspersing the writing with incongruous low-level mistakes in written English?
- Accounting for any differences in writing style necessitated by different assignment types/briefs, are there differences not thus accounted for that indicate a third-party involvement?
- Are gender and other individual personal attributes, explicit or implicit identity references consistent with the student's normal usage?

- An assignment that contains much narrative reporting, possibly with repetition, that does not align with the assignment focus.
- Otherwise good English written with circumlocutions where standard forms of words are expected/acceptable can indicate a ghost-writer writing to fulfil a 'plagiarism free' guarantee and subject-obliviously editing-out acceptable 'standard' forms of words that 'everyone' uses to reduce headline similarity.

2.3 References

- references that are listed but not cited – systematic, not occasional / accidental 'odd one or two';
- references that are cited but not listed – systematic, not occasional / accidental 'odd one or two';
- references that are not relevant to the assignment but have, for example, a keyword in the title that aligns with an assignment keyword;
- references that do not exist,
  - revealed as unhighlighted references in a bibliography-included similarity report (the expectation being that genuine references will have been cited previously in sources and are thus highlighted, implying unhighlighted references are probably not genuine).
- 'scattershot references', i.e. references that might – or might not – be (broadly) relevant, or which might not exist (see points above), which are cited at positions in the text where different points are being made;

2.4 Easter egg

- incongruous feature, e.g. an incongruous reference or statement that couldn't reasonably have been included by the student or missed when proof-reading etc., implying inclusion is deliberate to catch reader's / grader's attention.

2.5 Stylometric evidence – comparative analysis of style(s)

- This is seriously time-consuming, not straightforward to obtain, and requires statistical expertise to interpret.

- Significant stylistic differences between the suspected ghost-written assignment(s) with other submissions from the student, e.g.
  - aspects of style, vocabulary, use-of-language
  - that are unlikely to arise from the student's preparation of the assignment(s) in question,
  - aspects of style not indicated by assignment briefs,
  - and are more likely, on balance of probabilities to align with different authorship.

## 3    Tell-tales of AI-generated text.

3.1    Indicators that an assignment may have been 'written' by an AI text generator, include:

3.1.1  Headline

- Inconsistent writing style throughout the assignment, e.g. varying from sentence to sentence, paragraph to paragraph.
  - in particular ChatGPT and other AI text generators can stitch together strings of text from different sources with different styles, different language/grammar/punctuation conventions.
- Well-written English (or other language) but verbose containing hyperbole, circumlocutions, absences of insight, non-sequiturs, out-of-sequence reference points, inconsistent numerical ranges and comparisons.
- Absence of typos and other 'low-level' errors common in student-written (and, more generally, other human-written) text.

3.1.2  In more detail

- An assignment characterised by:
  - loosely connected paragraphs, lacking flow with variably phrased repetition of content over successive (pseudo-introductory) paragraphs;
  - grammatically correct but padded sentences containing subject-inconsistent, subject-disconnected clauses;
  - uniformity of sentence and paragraph styles;
  - sentences and paragraphs determined 'algorithmically' rather than according to development of subject.
- Examples of AI (GPT-3) mistakes (hallucinations), inconsistencies, hyperbole, circumlocution
  - *UK unemployment rose from 5% to 5% after the 2008 crash…*
  - *Leading climate scientists concluded that the earth warms in the temperature range of 7 degrees to 11 degrees Fahrenheit (1990-200 degrees Celsius) between 1990 and 2100.*
    - [NB. Interval (range) of 7–11 °F implies 3.9–6.1 °C]
  - *Go back to 2012 and Le Pen's message was essentially the same as it is today: a salty brew of jeremiads against the "Europe of Brussels," globalization, and the French political establishment.*

- *Workers are the backbone of the organization and his reactions to other people and the environment vary greatly depending on the value and principle he follows according to his life.*

### 3.1.3 References

- References that are not relevant to the assignment but have, for example, a keyword in the title that aligns with an assignment keyword.
- References that do not exist (example of AI 'hallucination', see 3.3),
  - ChatGPT in particular is/was notorious for inventing/faking references stitched together from bits and pieces of actual references (it's what it's programmed to do),
  - revealed as unhighlighted references in a bibliography-included similarity reports (the expectation being that genuine references will have been cited previously in sources and are thus highlighted, implying unhighlighted references are probably not genuine).
- 'Scattershot references', i.e. references that might – or might not – be (broadly) relevant, or which might not exist (see points above), which are cited at positions in the text where different points are being made.

### 3.2 AI-text detection (and appendix A1)

- AI-text detectors are, and never can be, reliable enough to be the sole determining factors of the presence or absence of AI-generated text.
  - This is unavoidable and a consequence of the underpinning mathematics – essentially 'information entropy'.
  - In short, human-written text is typically relatively high entropy compared to AI-generated text which is typically relatively low entropy, but AIs are getting more human-like (in this context) all the time.
- Evidence from all AI-text detectors is unreliable and should not be used as primary evidence.
  - At most, use AI-text detector information as secondary/supporting evidence, and use with the proviso that such evidence could well be misleading or incorrect.
  - Where possible, use more than one AI detector and, even if they agree, be aware that all could well be misleading or incorrect
- AI-text detectors can be biased: these tools determine 'probable AI' vs. 'probable human' according to their programming and their training data. Thus, for example:
  - if the training data contain disproportionately high levels of scholarly writing by confident native English users, whether students or academics, then the determination of 'probably AI' vs. 'probably human' will have that bias factored in and thereby disfavour non-native English users and non-scholarly writing styles;
  - students writing in English as their second, third... language can use straightforward, uncomplicated sentence constructions and vocabularies as can be typical of the 'low entropy' text an AI-text detector is programmed to detect as 'probably AI'.
- When using an AI detector, ensure that the text is appropriately redacted of any/all student name/ID information.

- AI-text detectors do <u>not</u> identify AI hallucinations *per se*: an AI-text detector will only identify an AI hallucination where the nature of the wording (i.e. the entropy, not the meaning, of the words) indicates 'probably AI' rather than 'probably human'.

3.3    AI Hallucinations (and appendices A2, A3)

- AIs are programmed to generate runs of words from user-entered prompts and questions, according to the probability that words / phrases / sentences occur in their databases. On occasion, depending on the prompt or question, an AI can 'get it wrong' and generate a section of text that is not relevant, not what was intended.

- An AI is not programmed to check the text it generates for veracity: in effect, it infers veracity from probability, on the assumption that a commonly-occurring run-of-words in its dataset is correct.

- irrespective of whether falsified/fabricated information results from an AI or a human, falsification/fabrication is serious academic misconduct.

- Typical AI hallucinations include, but are not limited to, the following.

  - Faked data/information, including:

    - real data/information in incorrect contexts and combinations,

    - factually correct but inconsistent data/information in same sentence, adjacent/nearby sentences.

  - Faked references.

    - In making a citation, a student is effectively declaring that they have read sufficient of the cited material to determine it supports their assignment. If the cited material doesn't exist, it can't support the assignment and the student can't have read it. Thus, faked references are as serious as other falsified/fabricated content.

  - Faked quotations.

    - In making a quotation, a student is effectively declaring that they have read the material they are quoting. If the quotation is faked, it can't support the assignment. Thus, faked quotations are as serious as other falsified/fabricated content.

- As of 27/09/2023, there is the additional possibility for ChatGPT, and other AIs that are programmed to search the Internet for up-to-date real-time information, that they 'hallucinate' data/information that post-dates their training datasets due to differences in context.

  - This could include rapidly-circulating 'viral' mis-/dis-information.

**Appendix: ChatGPT & Claude-2 generated answers to questions:**

**A1 Is AI-text detection biased?**

**A2 What is an AI hallucination?**

**A3 How can I identify AI text hallucinations?**

**A4 What are the tell-tales of AI-generated sentences and paragraphs?**

**A1  Is AI-text detection biased?**

A1.1  This is how ChatGPT replied to the question 'Is AI-text detection biased?' on 11/07/2023, verbatim.

- *AI-text detection systems can exhibit biases, just like any other machine learning model. These biases can arise from several sources, including biased training data, biased feature selection, or biased decision-making processes. If the training data used to train the AI-text detection system contains biases, the system can learn and amplify those biases in its predictions and classifications.*

- *One common issue is that AI models can reflect and reinforce societal biases present in the training data. For example, if the training data primarily consists of texts from a particular demographic or reflects discriminatory views, the model may learn to make biased decisions based on race, gender, or other protected characteristics.*

- *Moreover, the design and implementation of AI models, including the choice of features and algorithms, can also introduce bias. Biased feature selection or algorithmic decision-making can lead to unequal treatment or inaccurate predictions for certain groups or types of text.*

- *It is important to recognize that biases in AI-text detection systems are not inherent to AI technology itself but rather a reflection of the data and processes used to train and deploy these systems. Addressing bias requires careful consideration of the data used for training, ongoing monitoring and evaluation of the system's performance, and the implementation of appropriate safeguards and mitigation strategies to ensure fairness and accountability.*

A1.2  This is how Claude-2 replied to the question 'Is AI-text detection biased?' on 12/07/2023, verbatim.

- *AI text detection systems can absolutely exhibit biases, like any artificial intelligence system trained on human-generated data. Some potential issues with bias in AI text detection include:*

  - *Gender bias - Text detection systems may be more accurate at recognizing text written by or about certain genders due to imbalances in the training data.*

  - *Racial bias - Similarly, racial bias in datasets can lead to differences in accuracy for detecting text in different languages, written by authors with certain ethnic backgrounds.*

  - *Topic bias - Text detection systems may be "better" at recognizing text on certain topics they were more heavily trained on.*

- *Contextual bias - The context and format of text impacts detection. Systems may skew towards certain types of documents and styles of writing.*

- *Default bias - There can also be default assumptions built into the AI, e.g. assuming text is in a Latin alphabet unless otherwise specified.*

- *To reduce the potential for biased results, companies and researchers creating text detection systems need to carefully audit their training data, test performance across different demographic groups, and proactively sample diverse text. But like most AI, these systems are imperfect and reflecting the imbalanced data used to develop them. Being aware of the limitations and potential biases is important. Overall, text detection should not be viewed as universally objective and neutral - context matters.*

## A2    What is an AI hallucination?

A2.1    This is how ChatGPT replied to the question 'What is an AI hallucination?' on 04/07/2023, verbatim.

- *An AI hallucination refers to a phenomenon where an artificial intelligence system generates or perceives information that is not based on reality or actual data. It involves the generation of synthetic content by an AI system that appears realistic but is not grounded in actual observations or experiences.*

- *AI hallucinations can occur in various domains, such as computer vision, natural language processing, and generative modeling. For example, in computer vision, an AI system may generate images that resemble objects or scenes but are entirely fabricated. In natural language processing, an AI model might generate text that seems coherent but lacks factual accuracy.*

- *These hallucinations can arise due to limitations or biases in the training data, deficiencies in the model architecture, or errors in the learning process. AI systems rely on patterns and examples from their training data to make predictions or generate content, and if the training data is incomplete or biased, the AI may produce hallucinatory outputs.*

- *It is important to note that AI hallucinations are not intentional fabrications but rather unintended consequences of the AI learning process. Researchers and developers are continuously working to improve AI systems and mitigate such hallucinatory effects by enhancing data quality, refining models, and implementing robust validation mechanisms.*

A2.2    This is how Claude-2 replied to the question 'What is an AI hallucination?' on 12/07/2023, verbatim.

- *An AI hallucination refers to when an artificial intelligence system generates imagery or text that does not correspond to reality. Some examples of how AI hallucinations can occur:*

  - *Neural networks generate convincing images that depict objects, people or scenes that do not exist. This can occur when sampling randomly from the latent space of certain generative models like GANs.*

  - *AI systems can fabricate textual content, like news articles or reports, that sound plausible but describe fictional events. This demonstrates issues with veracity and fact checking in language models.*

- *Reinforcement learning agents may interpret randomness or noise in their environments as real sensory input, causing them to "hallucinate" obstacles, rewards, etc. that aren't actually there.*

- *Bugs in data processing can lead machine learning systems to "see" patterns that aren't real. Adversarial examples and other data perturbations can also cause this.*

- *Predictive algorithms can make highly incorrect extrapolations far outside their training distribution, seeing relationships that don't correspond to reality.*

- *The common theme is the AI perceiving or generating content that does not reflect the real world. Hallucinations demonstrate limitations in an AI system's ability to accurately model and represent the complexities of the real world. They show the potential divergence between data patterns and actual truth.*

## A3   How can I identify AI text hallucinations?

A3.1   This is how ChatGPT replied to the question 'How can I identify AI text hallucinations?' on 04/07/2023, verbatim.

- *Identifying AI text hallucinations can be challenging, as they often mimic the style and coherence of genuine text. However, there are several indicators that can help you detect potential AI-generated content. Here are some tips:*

  1. *Incoherent or nonsensical responses: AI models might generate text that sounds plausible at first glance but lacks logical coherence upon closer inspection. Look for inconsistencies, irrelevant information, or responses that do not directly address the context or question.*

  2. *Unusual or incorrect facts: AI models may generate information that is factually incorrect or improbable. If you encounter statements that seem dubious or contradict established knowledge, it could be a sign of hallucinatory text.*

  3. *Lack of context or specificity: AI-generated responses might lack contextual understanding or provide generic answers that do not directly address the given question. If the text feels generic and does not provide specific details, it could be an indication of AI-generated content.*

  4. *Overuse of jargon or technical terms: AI models trained on technical or specialized text sources may excessively employ jargon or use technical terms inappropriately. If you notice an overabundance of specialized vocabulary that seems out of place or unnecessarily complex, it could be a red flag.*

  5. *Repetitive or formulaic patterns: AI models can sometimes generate text that is repetitive or formulaic, especially if they have been trained on large quantities of similar content. If you observe a pattern of similar phrases or responses, it might indicate AI-generated content.*

  6. *Lack of common sense or human-like understanding: AI models often struggle with common sense reasoning and may generate responses that lack human-like understanding or intuition. If the text seems detached from everyday knowledge or lacks intuitive judgment, it could be an indication of AI-generated content.*

- *Remember that these indicators are not foolproof, and AI models are constantly improving. It's important to approach AI-generated text with critical thinking and skepticism, especially in situations where accuracy and reliability are crucial.*

A3.2  This is how Claude-2 replied to the question 'How can I identify AI text hallucinations?' on 12/07/2023, verbatim.

- *Here are some tips for identifying potential hallucinations or fabrications in text generated by AI:*

  - *Check factual accuracy - Verify claims against known facts from trustworthy sources. Watch for made up statistics, false historical details, etc.*

  - *Look for logical inconsistencies or contradictions within the text itself. Does it maintain internal coherence?*

  - *Examine the level of detail - AI text may use vague generalities when specifics would be expected. Lack of concrete details could indicate fabrication.*

  - *Evaluate the tone and style - Does it seem appropriate for the supposed author and context or is the writing oddly disjointed?*

  - *Watch for wandering off topic and tangents that don't make sense for the subject. Hallucinations may not stay focused.*

  - *Consider how well opinions, stances and conclusions are supported by evidence. Lack of sound support raises red flags.*

  - *Search for unique phrases or sentences in quotation marks to see if they appear elsewhere as AI can mashup or remix content.*

  - *Assess if the text meets your purpose - does it adequately answer your questions and information needs?*

  - *See if the AI can provide useful citations or source references if asked. Lack of sources may indicate pure fabrication.*

- *The key is applying critical reading skills and not taking AI text wholly at face value without verification. A degree of skepticism is warranted to catch potential hallucinations.*

## A4  What are the tell-tales of AI-generated sentences and paragraphs?

A4.1  This is how ChatGPT replied to the question 'What are the tell-tales of AI-generated sentences and paragraphs?' on 11/10/2023, verbatim.

- *Detecting AI-generated sentences and paragraphs can be more challenging because advanced AI models are designed to produce content that closely mimics human writing. However, there are some tell-tale signs that can help you identify AI-generated text at the sentence and paragraph level:*

  1. *Consistent Formal Tone: AI-generated text often maintains a consistent, formal tone regardless of the subject matter or context. This can make it sound slightly robotic or less natural.*

  2. *Lack of Personal Experience or Emotion: AI-generated sentences and paragraphs typically lack personal anecdotes, experiences, or emotional*

*nuances, as they are generated based on patterns in training data and lack personal consciousness.*

3. *Contextual Inconsistency: AI models may produce text that is contextually inconsistent within a sentence or paragraph. The content might abruptly change topics or use conflicting information.*

4. *Repetition: AI-generated content can exhibit repetition of words, phrases, or ideas within a sentence or paragraph, making it sound less varied and engaging.*

5. *Overly Long Sentences or Paragraphs: AI-generated text may include lengthy and complex sentences or paragraphs that are overly verbose or difficult to follow.*

6. *Vague or Ambiguous Language: AI-generated sentences and paragraphs might use vague or ambiguous language that requires additional context to understand fully.*

7. *Unusual Phrasing: While AI models are generally proficient in grammar and syntax, they may use phrasing or word choices that seem slightly off or unnatural.*

8. *Lack of Personal Perspective: AI-generated text tends to present information in an objective manner and lacks a personal point of view or subjective perspective.*

9. *Inaccuracies or Outdated Information: AI-generated content may contain factual inaccuracies, outdated information, or incorrect references because it relies on patterns in its training data.*

10. *Generic Quotes or References: AI-generated text may include quotes, references, or citations, but they may be generic or overly relied upon without adding substantial original content.*

11. *Difficulty in Clarification: AI-generated text may struggle to provide clear explanations or justifications when asked to clarify or expand on a statement within the text.*

12. *Artificially Complex Language: AI models may sometimes use overly complex language or vocabulary, making the text sound unnatural or out of place.*

13. *Incomplete Information: AI-generated text may produce content that is vague or lacks comprehensive information, leaving readers with unanswered questions.*

14. *Consistency Across Similar Content: If you notice multiple sentences or paragraphs with extremely similar style, structure, or information, it could be a sign that AI has generated them.*

- *It's important to keep in mind that the ability of AI models to mimic human writing is continually improving. As a result, the tell-tale signs mentioned above may become less evident in more advanced models. Always exercise critical thinking and consider the context when evaluating the authenticity of text, especially in situations where accuracy and credibility are essential.*

A4.2 This is how Claude-2 replied to the question 'What are the tell-tales of AI-generated sentences and paragraphs?' on 11/10/2023, verbatim.

- *Here are some common signs that text may have been generated by AI:*

- *Lack of consistent personality or voice - AI models aim for neutral, generic language without a distinctive authorial style. Human writing tends to have more variation and quirks.*

- *Repetition of words or phrases - AI models often repeat phrases or reuse the same vocabulary in unnatural ways. Human writing has more variation.*

- *Disjointed logical flow - AI text may have sentences that don't clearly connect or transition between ideas. The flow of logic can be disjointed.*

- *Lack of concrete details - AI text often lacks specific names, places, dates etc and uses more general/abstract language. Human writing is more likely to have concrete details grounded in reality.*

- *Grammatical errors or inconsistencies - Since AI models don't have a full understanding of grammar rules, they can make small errors or be grammatically inconsistent.*

- *Unnatural responses - When used for dialogue, AI systems can respond generically or say things that don't make sense in context. Human dialogue aligns better with the flow of conversation.*

- *Irrelevant text - AI models sometimes include text that is tangential or completely irrelevant to the prompt or context. Human writing stays on topic.*

- *So in summary, a lack of voice, repetition, logic gaps, generic text, errors, unnatural responses, and irrelevant text are red flags that prose may be AI-generated. Looking at the overall flow and consistency can reveal if something was written by a human or machine.*

References re. Appendix:

https://chat.openai.com/auth/login, 04-11/07/2023 & 11/10/2023

https://www.anthropic.com/index/claude-2 12/07/2023 & 11/10/2023

2023-24 v1

23/11/2023