

A Comparison of Amplifi and Microsoft Copilot in Word: Simplifying Regulated Communications for Compliance

Executive Summary

This study compared the performance of Amplifi (a specialist reg-tech tool) and Microsoft Copilot (in MS Word) in simplifying complex, regulated communications, specifically within the UK financial services sector.

Main Findings

The study concludes that the primary difference between the two tools is systemic: Amplifi functions as an operational risk control, while MS Copilot acts as a generative drafting assistant.

- **Measurable Intelligibility:** Amplifi produced significantly higher gains in document intelligibility. In paragraph-by-paragraph testing, Amplifi improved a credit card agreement's score by 66%, compared to MS Copilot's 31%.
- **Diagnostic vs. Generative:** Amplifi identifies specific complexity drivers prior to editing. MS Copilot lacks native diagnostics, forcing users to rely on subjective judgment to find risks.
- **The Scoring Inflation Risk:** A critical finding revealed that MS Copilot significantly over-scored its own work. In one instance, MS Copilot gave a simplified press release a score of 82/100, while Amplifi's objective assessment using its validated method was only 50/100. This creates a material compliance risk by providing false evidence of clarity.

Impact on Compliance and Governance

The study highlights several potentially high stakes impacts for firms managing Consumer Duty and other regulatory obligations:

1. **Semantic Drift:** MS Copilot frequently weakened precise legal qualifiers or altered the sequencing of terms. This introduces hidden risk where a document becomes easier to read but loses its legal integrity.
2. **Auditability:** Amplifi provides a structured journey from initial assessment, to editing, and then reporting. It generates a traceable audit trail suitable for senior management and regulators. MS Copilot lacks an integrated governance architecture, shifting the burden of validation entirely back to the human user.
3. **Efficiency vs. Effectiveness:** While MS Copilot accelerates the *speed* of drafting, it does not reduce the *diagnostic burden*. Amplifi reduces the total effort by directing the user to high-risk clauses, ensuring interventions are high-impact and evidence-led.

Conclusion

For regulated environments, the study suggests that Amplifi is a risk-managed system that provides defensible proof of improvement. Conversely, MS Copilot is a more unstable environment for sensitive text, where performance is highly dependent on individual prompt engineering and manual human oversight to prevent semantic drift.

MS Copilot's limitations in text simplification and assessment

High dependence on prompt wording

Simplification is contingent on how the task is phrased, which introduces user variability into the process.

Poor control over editing scope

MS Copilot does not always act only on the selected or intended text, increasing the risk of unintended edits and weakening trust in the editing workflow.

MS Copilot's interaction model is inconsistent

Sometimes allowing the user to review alternative text before applying it, at other times it writes directly into the document. This reduces transparency and control.

Inconsistency across repeated runs

Similar prompts can produce materially different outputs, both in content and in how the system applies the task.

Weak reversibility and unstable editing recovery

Restoring or reviewing the original wording is not always straightforward. This shifts the burden back onto manual document management and version control.

MS Copilot takes initiative beyond the user's explicit instruction

The system not only generates language but sometimes redefines the task itself. That increases the likelihood of unintended changes.

MS Copilot's assessment of intelligibility is not reliable

It tends to produce scores that are materially more favourable. It cannot be relied upon to validate the success of its own simplification.

Risk of semantic and structural drift

Legally significant wording was frequently softened, qualifiers weakened, and sequencing was altered. This required manual checking, reducing efficiency and increasing risk.

Governance and auditability limitations

Prompt history, output history, and revision logic are not always captured in a way that supports robust audit trails. Governance controls must be recreated externally.

Report

Aim of the study

This assessment tested the extent to which different AI approaches can support the safe simplification of regulated, legally sensitive communications for a general audience while preserving legal meaning, mandatory disclosures, and compliance integrity.

The study compares not only final rewrites, but the full operating model: how each tool diagnoses complexity, guides intervention, supports user decision-making, and generates evidence suitable for internal governance and regulator-facing assurance.

The core objective is to quantify where each tool reduces effort without creating hidden risk via omission, drift, or weak auditability.

Context

Intelligibility is a conduct risk and a factor in creating consumer harm. It can be hard to measure as factors leading to misunderstanding are many and varied and often scattered throughout a document.

These factors can include specific structural features, such as:

- conditional stacks (“if... unless... except when...”)
- cross-references
- defined terms which may be unusual or technically complex
- embedded exceptions
- legally loaded qualifiers (“including but not limited to”, “subject to”, “may”, “reserve the right”).

These features can create comprehension failure even when the vocabulary is simple.

Simplification that primarily targets “readability” can create a dangerous illusion of clarity. In regulated text, the issues that matter can include ambiguity, weakening a qualifier, dropping a required disclosure, changing sequencing (what applies when), or introducing a friendly but false commitment.

This is why simplification of key regulated information must work as a controlled tailoring process, rather than just a stylistic rewrite.

Controlled tailoring requires three things:

- a measurable baseline and target,
- a method for identifying and prioritising high-risk complexity,
- a verification mechanism that demonstrates that all the necessary meaning and disclosure completeness were preserved.

The fundamental question is therefore whether a tool can act as a risk-managed simplification system, not whether it can draft plain English.

What we assessed, and how

To undertake this assessment, we compared two ways to utilise AI and LLMs to assess and simplify regulated information. In this case consumer-facing financial service documents. We would like to widen this assessment to other document types and sectors in future.

To do this, we compared the results using an established and widely available generalised LLM, Microsoft Copilot, with a specialist intelligibility assessment and simplification tool, Amplifi.

About Amplifi

Amplifi is a specialist reg-tech tool designed to assess and improve intelligibility using a wide range of measurable, objective metrics proven to accurately predict understanding outcomes. It tests written information, providing an intelligibility score and estimated educational equivalent and reach scores. It uses a diagnostic workflow to identify specific structural drivers of complexity, allowing users to tailor information where comprehension risk is highest using targeted suggestions and continual intelligibility assessment. Academic and regulator-partnered testing has ensured that it is a robust method for measurement-led compliance. Its current operational limitations include formatting-related challenges, such as consistently integrating structured content like tables and numerical formulae into its automated simplification and scoring processes.

About Microsoft Copilot

Microsoft Copilot in Word functions as a general-purpose generative drafting assistant that excels at rapidly rewriting text into plain English and maintaining document formatting. Its primary strength is reducing the friction of content generation by quickly producing alternatives, such as FAQs or summaries, and handling structured outputs like bulleted steps or tables. However, because it is a multi-purpose LLM, it lacks native, tool-surfaced intelligibility diagnostics and does not provide an objective measure of whether a rewrite actually reduces comprehension risk or increases audience reach. Consequently, while it accelerates the initial drafting stage, it shifts the critical burden of validation and governance onto human users.

Both of these tools and other generalised LLMs are being widely used by financial service firms, and specifically content creators and governance specialists, to test and tailor (simplify) documents to comply with the Consumer Duty. We also tested the capacity of these tools to report on the work undertaken as part of a firm's evidence of compliance and internal and external reporting.

Assessment Framework

Focusing primarily on the risk outcome, we tested the two technologies across a typical journey: Initial Assessment -> Simplification -> Re-assessment -> Reporting.

Following conclusion of the process, we measured the success of the two tools across four risk-related layers:

1. **Intelligibility Risk and Reach:** Did the tools reduce risk and increase likely comprehension, and did they make the drivers of complexity visible?
2. **Compliance-safe Simplification:** Did the tailoring of the document retain the semantic fidelity and completeness of the disclosures?
3. **Governance, Auditability and Scale:** Can the tools help users prioritise a portfolio, apply targets consistently, and produce evidence suitable for sign-off and regulatory reporting?
4. **User Journey and Efficiency:** Did the tools reduce total effort from intake to approved output, without shifting hidden validation work back onto the user?

We used three simplification strategies to expose tool behaviour under progressively more realistic constraints:

- Paragraph-by-paragraph single-prompt strategy to test core capability without prompt engineering.
- Hybrid strategy to test a more realistic regulated workflow: diagnose complexity, apply targeted prompts, iterate, and validate.
- Suggestion-led strategy to test whether the tool can drive editing based on the complexity it identified, rather than led by user prompting.

We tested the tools using these common strategies across a number of publicly available whole documents, and document excerpts:

- Nationwide Credit Card Agreement – Minimum Payments Section
- Ocean Credit Card Agreement (Full Agreement)
- Intellectual Property Rights Section (Terms of Use)
- Standard Life Privacy Notice (Selected Sections, including profiling and decision-making)
- Standard Life Press Release (Consumer-facing financial communication)
- Standard Life Funds and Charges Document (Lifestyle profiles section)

The assessment intentionally included both high-risk legal content (such as agreements and disclosures) and lower-risk but still complex communications (such as press releases). This ensured that the findings reflect not only worst-case regulatory scenarios but also everyday communications.

Results and insights

Paragraph-by-paragraph simplification using a single prompt

Document	Original Score	Amplifi Paragraph Strategy	MS Copilot Paragraph Strategy
Nationwide Minimum Payments	68	75 +10%	71 +4%
Ocean Credit Card Agreement	35	58 +66%	44 +31%

Applying a single consistent prompt paragraph by paragraph is an efficient way to simplify a long and complex text.

The Amplifi tool, using a consistent, simple prompt had a much more significant improvement in the intelligibility of the text for the same time spent.

Amplifi reduced cognitive guesswork. The tool shows whether the suggested simplified text has improved intelligibility during the editing process at each step. Compared to generative rewriting systems, Amplifi introduces discipline into the process.

MS Copilot on the other hand provides no objective signal that intelligibility has actually improved. This forces users to rely entirely on subjective judgment. In dense clauses, it occasionally replaced precise legal terms with broader wording, potentially weakening the accuracy of triggers or qualifiers, a factor which was largely absent using Amplifi.

Ultimately, while MS Copilot accelerates drafting, it does not reduce the diagnostic burden. Risk detection remains a manual process that depends heavily on the user's own legal literacy and attentiveness.

Hybrid workflow combining prompting with complexity guidance

Document	Original Score	Amplifi Hybrid Strategy	MS Copilot Hybrid Strategy
Intellectual Property Rights	55	72 +31%	61 +11%

The hybrid strategy demonstrated Amplifi's governance architecture. When users can alternate between identifying complexity and targeted simplification, the

workflow becomes evidence-led rather than exploratory. Users can prioritise high-risk clauses identified by the system and verify improvement after each iteration.

Amplifi changes the editing sequence. Instead of drafting first and validating later, users diagnose first, intervene in a targeted way, and measure the impact of their work continuously. This produces larger intelligibility gains in dense, complex text because effort is concentrated where it has the highest impact.

Importantly, the hybrid workflow demonstrated that shorter, more targeted prompts guided by diagnostics in a dedicated tool will outperform long constraint-heavy prompts attempting to achieve the same ends in a general LLM. This reduces prompt dependence and risk. Amplifi's embedded signals do more of the governance work than prompt engineering can reliably replicate.

The hybrid strategy is where MS Copilot's prompt dependence becomes most visible. Because MS Copilot does not natively surface complexity drivers, users had to prompt the system to identify difficult areas before simplifying them. The quality and completeness of that diagnosis depended heavily on how the question was phrased.

When provided with long, constraint-heavy prompts such as instructions to retain capitalised terms, preserve format, and limit sentence length, MS Copilot sometimes followed the instructions but not consistently. In some cases, defined terms were altered, sequencing was affected, and formatting constraints were partially ignored.

This highlights an architectural difference and key limitation of MS Copilot: governance instructions included in prompts are not equivalent to embedded system controls. Users must attempt to 'instruct' compliance safeguards into the model rather than rely on system-level enforcement. This creates variability and introduces risk.

In more iterative shorter prompt cycles, MS Copilot sometimes referenced earlier versions of text rather than the latest edited draft, creating ambiguity about what had changed. This introduces challenges for auditing the work done, because the relationship between input, suggestion, and output is not systematically tracked.

However, Amplifi was not without some shortcomings. The hybrid simplification strategy also revealed some constraints that can cap performance. Score improvements are not guaranteed after every edit, and some iteration is necessary based on the feedback it provides.

The key insight here is strategic: Amplifi performs best when used to identify, measure and prioritise risk, to guide human intervention. Its primary differentiation lies in a more certain workflow and risk identification.

Suggestion-led editing based on tool-identified complexity

Amplifi's suggestion-led approach provides a significant conceptual advantage by surfacing complexity immediately, allowing editing to begin with diagnosis rather than manual prompting. This shifts user behaviour from reactive rewriting to targeted risk reduction.

By making complexity drivers visible, users can intervene at the clause level, which improves precision, reduces the risk of omitting key information, and ensures the workflow remains traceable for regulatory governance.

MS Copilot’s significant limitation in suggestion-led editing is its inability to automatically identify complexity drivers, requiring users to explicitly request help. This results in guidance that is often stylistic rather than structurally analytical, and fails to address the root causes of poor understanding,

Ultimately, because MS Copilot responds to prompts rather than embedded diagnostics, simplification is not driven by structured risk prioritisation. This leaves the user entirely responsible for deciding where to intervene, making performance highly dependent on individual skill and attention. This reduces the consistency and efficacy of the results.

Assessing intelligibility and likely understanding outcomes

Amplifi has been developed specifically with Intelligibility assessment in mind: i.e., the ability to accurately and objectively test real life likelihood of understanding outcomes. This has been tested and validated with both the FCA and the Solicitors Regulation Authority, and with significant academic input from a range of leading academics.

We scored the original unedited version of a Standard Life press release with Amplifi and it scored just 41 out of 100. This represents the equivalent of an undergraduate level of education required to fully understand and apply the information, achieving a reach of around a third of UK adults.

After simplifying it using MS Copilot, we then used both Amplifi and MS Copilot to re-assess it for intelligibility. We prompted MS Copilot to assess the intelligibility of using a scale from 1-100, similar to the Amplifi Intelligibility Score.

Example: Standard Life Press Release (Consumer Communication)	
Amplifi Intelligibility Score	MS Copilot Intelligibility Score
41	82

Worryingly, it significantly over-scored its work, suggesting a score of 82 out of 100. This would provide editors with false evidence of intelligibility, and presents a material risk for compliance, and potential for consumer harm if it were applied to consumer-facing disclosures.

This was tested on a number of other documents, the results of which are shown below.

Document	Original Amplifi Score	MS Copilot Simplified Version		Observations
		Amplifi Score	MS Copilot Score	
Standard Life - press release	41	50	82	MS Copilot shortens sentences but output remains weak; major scoring inflation
Standard life - privacy notice - section 9: Your Rights	66	68	92	Minimal real improvement despite very high self-score
Standard Life - Lifestyle profiles section - funds and charges	61	70	92	Moderate gain, but heavily inflated self-assessment

Comparing MS Copilot and Amplifi for task-based text simplification

Text simplification in regulated contexts is not a generic rewriting exercise. It is a task-based objective with explicit constraints:

- Reduce intelligibility risk
- Preserve legal meaning
- Retain all required disclosures
- Avoid introducing ambiguity or new commitments

This requires a system that can not only generate text, but undertake the task within defined boundaries, and verify that it has been completed correctly.

Copilot, as a generative AI system, is not designed to complete editing tasks within these requirements. It is designed to produce plausible language outputs in response to prompts. This creates a fundamental mismatch between the system architecture it is built around, and task required for regulatory or legal drafting.

By contrast, Amplifi approaches simplification as a measurable, controlled process. The task is defined, bounded, and evaluated through explicit metrics. It also encourages human intervention to verify and evaluate the outputs of the model.

Prompt engineering vs system control

In MS Copilot, as with other non-specialised LLMs, prompt engineering acts as a proxy for system design. Users must use self-generated prompts to encode task requirements into their instructions. This approach has inherent limitations.

1. Prompts are not enforceable constraints. They are interpreted probabilistically. The system may follow them, follow partially, inconsistently, or not at all. This leads to variability in output, even when the same prompt is used repeatedly.

2. Prompts must often carry multiple objectives simultaneously: simplify, preserve meaning, maintain format, avoid omissions. As prompt complexity increases, adherence typically decreases. This creates a trade-off between how clear the instruction is, and the reliability of its execution.
3. Prompt engineering introduces user dependency. Outcomes vary based on how well the user articulates the task. This makes the system difficult to standardise across teams and workflows.

A specific point solution such as Amplifi addresses these limitations by removing reliance on prompts as the primary control mechanism. While prompts can still be used, they operate within a structured environment where:

- Complexity is pre-identified
- Output targets are defined
- Progress is measured and recorded

This shifts control from user-defined instructions to system-level governance, reducing variability and improving consistency.

Conclusions

The fundamental distinction between Amplifi and MS Copilot is systemic rather than incremental.

Amplifi is built to diagnose and reduce intelligibility risk with measurable evidence, while MS Copilot focuses on generating and editing text. In regulated environments, Amplifi's model aligns with governance and audit obligations by shifting simplification from subjective rewriting to observable risk reduction. By surfacing complexity-driver diagnostics, it allows users to intervene where structural density is highest and provides defensible proof that edits improved consumer reach without compromising legal integrity.

Unlike general AI tools, Amplifi serves as an operational control within a communications risk framework. It enables portfolio-level prioritisation and evidence-based sign-off. The platform's measurement-led approach aligns directly with regulatory requirements to evidence improvement rather than relying on perceived clarity. While MS Copilot is a capable drafting accelerator for producing alternative phrasings, it lacks the diagnostic signals required to independently confirm that complexity has been reduced or that material disclosures remain intact.

Using a tool like MS Copilot shifts the diagnostic burden entirely to the user, who must manually assess whether meaning has drifted or legal qualifiers have been weakened. Amplifi provides direct feedback on whether a change actually improved intelligibility. This ensures a disciplined and traceable editing loop. For legally sensitive content, this measurable reduction of risk and the ability to demonstrate before-and-after changes are the critical differentiators for managing and evidencing intelligibility at scale.

Summary of differences			
Hybrid Workflow Stage	Amplifi	MS Copilot	Structural Consequence
1. Document Intake/upload	Document uploaded into a diagnostic environment that prepares it for analysis. This capability facilitates a substantial volume of uploads.	Document opened or pasted into drafting interface (MS Word). The functionality is limited to processing a single document.	Amplifi prepares for risk assessment. MS Copilot prepares for rewriting.
2. Baseline Assessment	Automatic generation of document and section-level intelligibility scores. Complexity drivers surfaced immediately.	No automatic scoring or structural analysis. User sees only text.	Amplifi establishes a measurable starting point. MS Copilot begins without a quantified baseline.
3. Complexity Identification	Amplifi highlights high-risk clauses based on structural drivers (conditional density, passive load, defined terms).	User must prompt: "Identify complex areas." Output varies by phrasing and scope.	Amplifi standardises diagnosis. MS Copilot introduces variability and prompt dependence.
4. Prioritisation	Sections ranked by risk severity. Editing guided by impact potential.	User manually decides where to start, often sequentially.	Amplifi optimises for risk impact. MS Copilot relies on user intuition.
5. Target Setting	Measurable intelligibility uplift target defined. Preservation constraints implicitly embedded in workflow.	User encodes constraints in prompt (retain meaning, keep format, limit sentence length).	Amplifi enforces structure via scoring and risk assessments. MS Copilot relies on linguistic instruction.
6. Clause Selection	User selects the highest-risk clause identified by the Amplifi tool or their judgement.	User highlights or copies chosen sections manually.	Amplifi intervention is risk-prioritised. MS Copilot intervention is workflow-driven.
7. Editing Action via AI Assistant	Short, targeted prompt applied to a specific structural issue.	Broader generative rewrite prompt applied. Often instruction-heavy.	Amplifi promotes precision. MS Copilot promotes fluency.
8. Structural Safeguard Check	User verifies preservation of qualifiers, disclosures, and defined terms before progressing.	User manually checks for omission, drift, or oversimplification.	In both tools human review exists, but Amplifi provides contextual diagnostics.

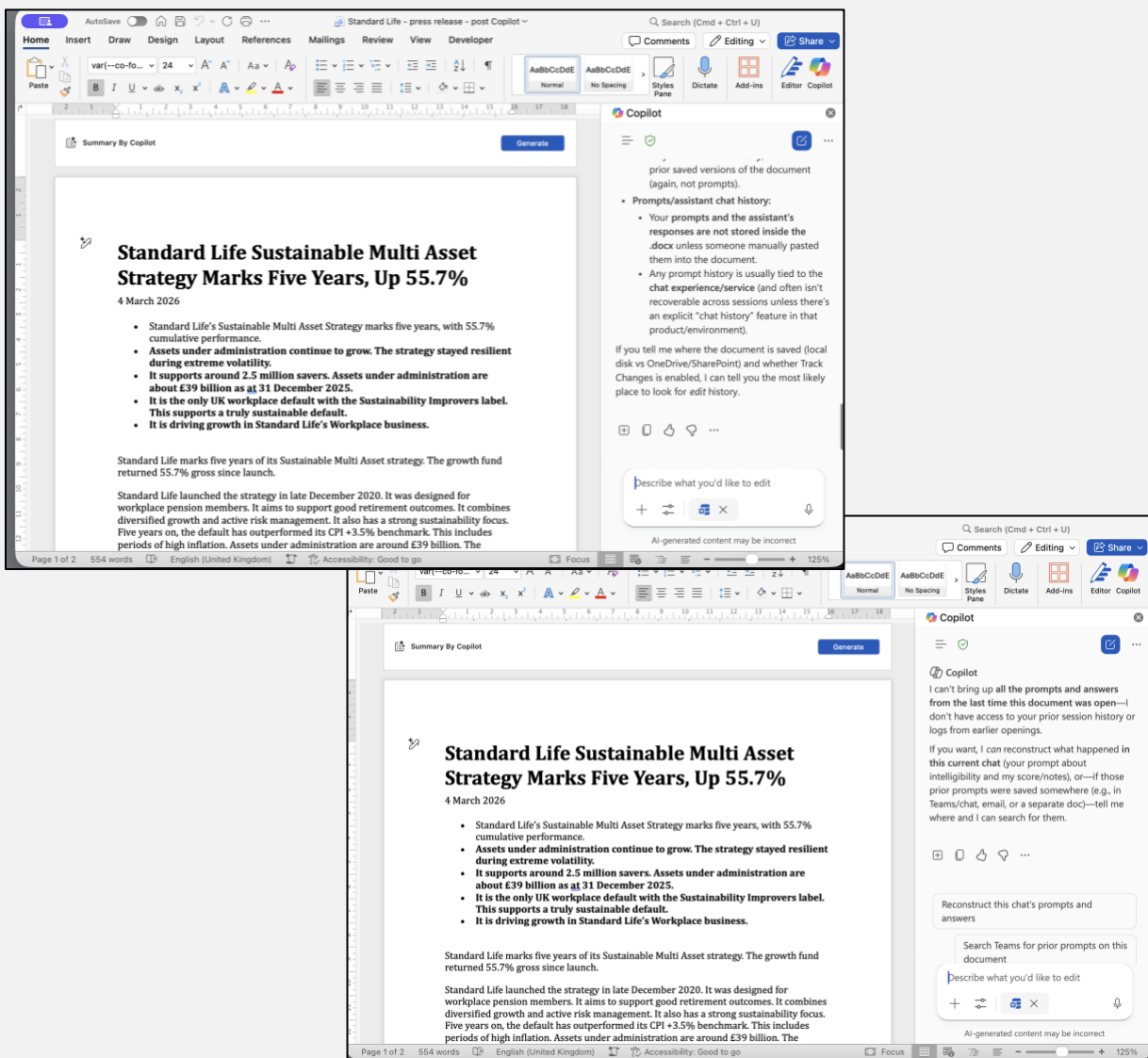
9.Re-Measurement	Section and document scores recalculated. Improvement quantified immediately.	No re-measurement. Improvement judged subjectively.	Amplifi closes the feedback loop. MS Copilot operates without quantitative validation.
10. Decision Gate	Continue only if intelligibility improves without semantic loss. If not, refine intervention.	Continue based on perceived clarity. No measurable threshold.	Amplifi provides objective stopping criteria and human judgement. MS Copilot relies on judgement.
11. Iteration Logic	Move to the next highest-risk clause based on ranking.	Continue rewriting sequentially or by user choice.	Amplifi maintains risk-priority discipline. MS Copilot lacks structured sequencing.
12. Document-Level Validation	History of edits is recorded and intelligibility score and reporting are provided.	Conduct manual read-through or track changes comparison.	Amplifi provides a report. MS Copilot requires manual reconciliation.
13. Governance Output	Produce before/after scores, documented improvement, traceable edit logic.	Governance achieved via external version control and documentation.	Amplifi generates audit artefacts. MS Copilot requires a separate governance process.

Examples of challenges we found

MS Copilot failed to provide an audit trail of its AI-driven simplification

When using MS Copilot to simplify text using a series of prompts, keeping a record of the changes made, the original text and the ability to use version control were limited beyond the initial session. This is a challenge when demonstrating the impact of tailoring information for the Consumer Duty, and to use MS Copilot as a content governance tool. The screenshots below were using MS Copilot to simplify an excerpt from a Standard Life document.

Amplifi has this functionality built into its operation as standard, as. It is designed as a governance and compliance tool.



Both tools failed to retain formatting accurately

1 Original unsimplified text

3.2 Joint Accounts. If this is a joint account, each of you will be individually and jointly responsible for paying all amounts owed under this Agreement. This means that we can require any one of you individually to repay the entire amount owed under this Agreement. Each of you authorizes the other(s) to make Purchases, Balance Transfers or Cash Advances individually. Any one of you may terminate the Account and the termination will be effective as to all of you. You must not allow anyone else access to the Card or the account number of the Account unless you authorize him or her to obtain credit under the Account. Except to the extent prohibited by applicable law, all credit obtained under your Account by anyone you allow to use your Card, or the account number of your Account will be covered by this Agreement as though you had obtained it. Consequently, you will be responsible for repayment of the debt.

2 MS Copilot simplified version

Joint Accounts. If this is a joint account, each person is fully responsible for the entire balance. We may require any one of you to repay the full amount owed. Each joint Holder allows the others to make Purchases, Balance Transfers, and Cash Advances. Any one of you may terminate the Account for all Holders. Do not let anyone use the Card or number unless you authorize them. You are responsible for credit obtained by anyone you allow to use the Account. You must repay all debt created by anyone you allowed to use the Card.

3 Amplifi simplified version

3.2 Joint Accounts

3.3

3.4 If you share this account with someone else, you are both responsible for all money owed. We can ask either person to pay the full amount.

3.5

3.6 Each person can make purchases, transfers, or cash advances alone. Any person can close the account for everyone.

3.7

3.8 Do not let anyone else use your card unless you authorize them. You are responsible for all money spent by anyone you allow to use your card.

Text 1, above, is a section of the Ocean Bank Credit Card terms and conditions, regarding joint account rules. The text was simplified as a whole paragraph in both MS Copilot (text 2) and using Amplifi (text 3) using an identical prompt: *"simplify this text. Reduce sentence length to 18 words or fewer"*

The MS Copilot version, while simplifying the text, has removed the numbered formatting and retained a lot of complexity, including the dense paragraph structure a lot of specific information.

Amplifi simplified the text further while retaining the key information. While it has broken down the dense paragraph, it has also created incorrect formatting, by numbering the separate lines.

MS Copilot significantly over-scored when asked to assess intelligibility

- 1 The right to have your personal information transferred to another organisation (the 'right to data portability') You can ask us to transfer any of your personal information to a specific company or person. We will try to transfer the information in the format you request. If this is not possible, we will contact you to agree an alternative format.

- 2 Transfer your personal information (right to data portability) You can ask us to transfer your personal information to a company or person you choose. We will try to use your requested format. If we cannot, we will agree another format with you.

- 3 The right to get your personal information deleted (the 'right to be forgotten')In certain circumstances you can ask us to stop keeping or using your personal information or ask us to restrict how we use it. If we can delete your information we will, but sometimes we must keep it for legal reasons. If we cannot meet your request, we will contact you to explain why.

- 4 Delete your personal information (right to be forgotten) In some cases, you can ask us to delete your personal information. You can also ask us to stop using it. Or you can ask us to restrict how we use it. We will delete your information when we can. Sometimes, we must keep it for legal reasons. If we cannot meet your request, we will explain why.

Text 1 and Text 3 are original sections of the Standard Life Privacy Notice. Text 2 and Text 4 are the versions of those text extracts simplified by MS Copilot using the same prompt.

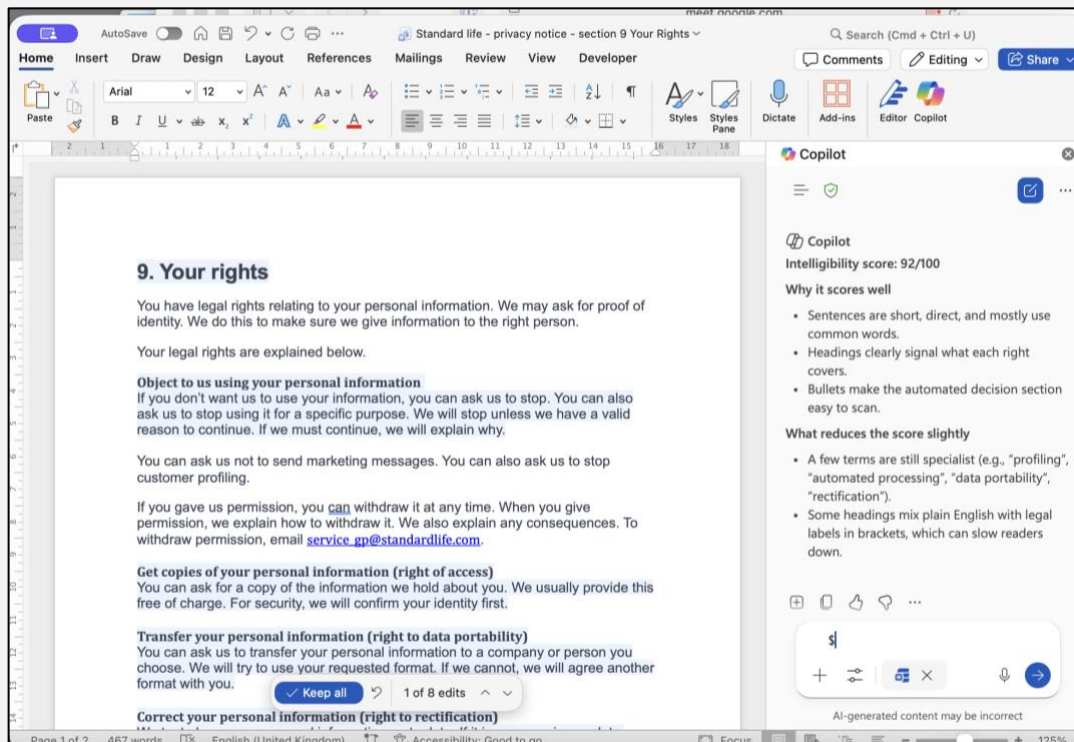
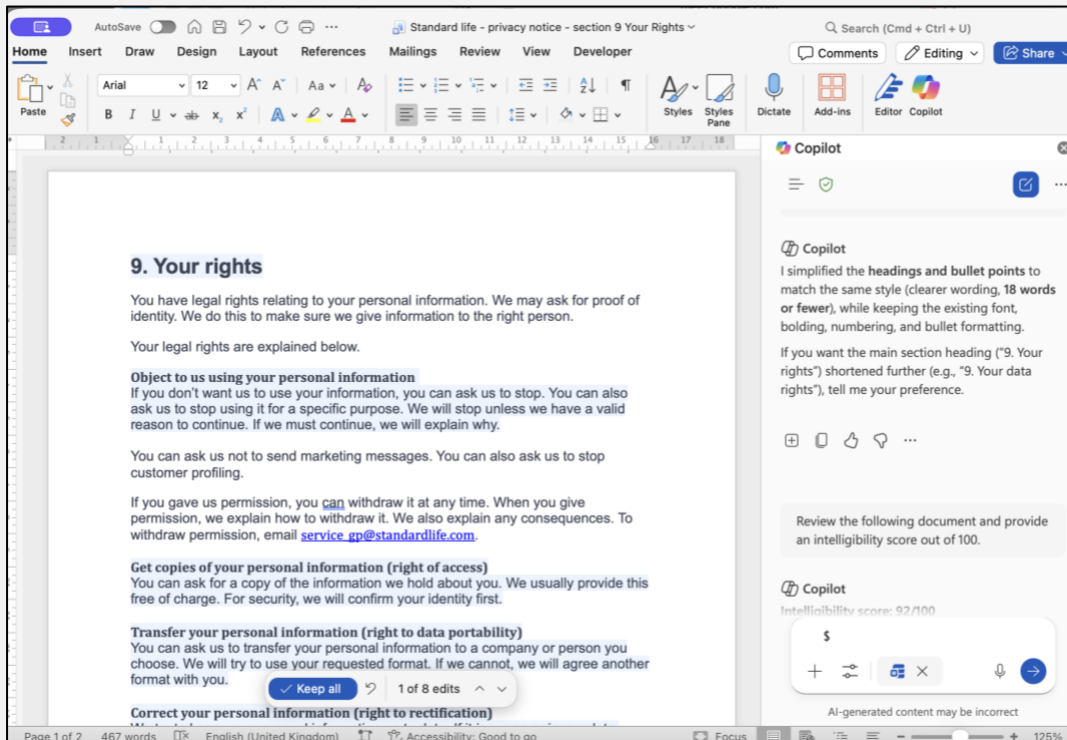
MS Copilot has simplified the text a little and retained much but not all of the necessary information (e.g. omitting "we will contact you"). And it has introduced some ambiguity in places.

The biggest challenge was when MS Copilot was asked to assess the Intelligibility of the document on a scale from 1 to 100.

The original was scored by Amplifi as a 66, and Copilot's slight simplification as a 68. However, MS Copilot assessed its own simplification to be transformative – scoring it 92 out of 100.

This fails to identify the true intelligibility and misleads the user to believe their document will be almost perfectly understandable and compliant.

The screenshots below show MS Copilot's results when asked to provide an intelligibility score for its simplified version of an excerpt from the Standard Life Privacy Policy (as also shown above).



MS Copilot lost more information when simplifying

1 **Original unsimplified text**

Verification of Information. You authorize an investigation of your credit standing and the verification of the information given in the credit card application, including your employment and income history, prior to the issuance of a Card to you or an Authorized User, and at any time thereafter. You also authorize disclosure of information to consumer reporting agencies, including but not limited to your late payments, missed payments or other defaults on your Account.

2 **MS Copilot simplified version**

Verification of Information. You allow us to check your credit and verify your application information at any time. You also allow us to report information about your Account to consumer reporting agencies.

3 **Amplifi simplified version**

Verification of Information. You allow us to check your credit history and verify information from your credit card application. This includes checking your job and income history before we give you a Card. We can also check this information anytime after that. You allow us to share information with credit reporting agencies. This includes late payments, missed payments, or other problems with your Account.

Text 1 is a section of the Ocean Bank Credit Card terms and conditions, relating to verifying the customer's information.

This text was then simplified as a whole paragraph in both MS Copilot (text 2) and using Amplifi (Text 3) using an identical prompt: *"simplify this text. Reduce sentence length to 18 words or fewer"*

The MS Copilot version, while simplifying the text, has removed a lot of specific and necessary information.

Amplifi simplified the text while retaining the key information.