Lessons Learned from Organizing a Literary Competition of Short Stories Co-created by Humans and AI Tools*

Rudolf Rosa
Charles University
Faculty of Mathematics and Physics
Praha, Czechia
rosa@ufal.mff.cuni.cz

Michaela Liegertová
Faculty of Science
Jan Evangelista Purkyně University
in Ústí nad Labem, Czechia
michaela.liegertova@ujep.cz

Miloslav Linc
Městská knihovna v Praze
(Municipal Library of Prague)
Praha, Czechia
miloslav.linc@mlp.cz

Abstract

In the past three years, we organized three rounds of a literary competition, in which human participants are asked to co-create short stories in collaboration with AI tools based on large language models. We collect, process and analyze the stories, ratings and comments by human judges, and participants' reports on their creation process. We gain various insights centered around several research questions: Which roles do humans and AI tools take on in the co-creative process? Where is human creative contribution necessary in the co-creation process (if anywhere) to produce high quality stories? What is the effect of literary and technical experience of the participants on the process and results? Do AI tools support or inhibit creativity?

Introduction

There recently has been increasing interest in AI-generated and co-generated stories and other literary texts (Teleki et al. 2025; Ismayilzada, Stevenson, and van der Plas 2025), with some results suggesting that Large Language Models (LLMs) can already be on par with humans or even outperform them (Gómez-Rodríguez and Williams 2023; Zhao et al. 2023; Porter and Machery 2024; Chakrabarty and Dhillon 2026) – especially when using the English language, for which the performance of LLMs is much stronger than for other languages (Zhang et al. 2023).

In our work, we investigate the state and development of capabilities of current LLMs for generation of short stories in Czech, which is a medium-resourced Slavic language. Over the past three years, we organized three rounds of a literary competition which required the participants to submit short stories co-created with the help of LLMs as well as potentially other AI tools. The stories were judged by experienced human judges, who produced rankings of the stories as well as some further insights. We combine these with reports about the collaboration process, provided by the competition participants, to analyze the co-creation processes and its relation to the resulting quality of the stories.

All data and resources that we are allowed to share can be found at our GitHub: `https://github.com/ufal/aiai`

*Research supported by Czech Science Foundation grant 25-14501L and by Charles University grant UNCE/24/SSH/009.

Research Questions

The main research questions we pose are:

1. Which roles do humans and AI tools take on in the co-creative process?
2. Where is human creative contribution necessary in the process (if anywhere)?
3. What is the effect of literary and technical experience of the humans?
4. Do AI tools support or inhibit human creativity?

Main Findings

A summary of our main findings is as follows:

- Both humans and AI tools can take on any of the roles in the creative process; but some scenarios are more typical than others.
- Human creative input is important *somewhere* in the process to achieve high-quality results; however, there is no single point in the process that would require human contribution.
- Too low as well as too high human involvement may be detrimental to the quality of the result; optimal results are achieved by a balanced human-AI collaboration.
- Experience in using AI tools and sophistication of the co-creation process are vital in producing a good output. However, this may get less important over time as the AI tools are significantly improving.
- Experienced writers did not outperform amateurs; writing experience does not seem to be helpful in the human-machine collaboration.
- AI tools can support the existing creative potential in amateur writers who lack some of the necessary writer skills. If used adequately, they do not kill the creative process but rather enable it.
- Experienced writers, capable of writing high-quality literature without AI help, were not observed to benefit from involving AI tools directly into the creative process; this rather disrupts the process and leads to worse results. However, they can benefit from AI tools for auxiliary tasks, such as research, proofreading, or critical feedback.

Round	Year	Focus	Participants
1	2023	sci-fi, fantasy, horror	74
2	2024	detective story, thriller	23
3	2025	explicit fictional author	53

Table 1: Rounds of the competition

The Competition

The Municipal Library of Prague has been working with aspiring authors for decades, providing them with various resources, offering creative writing workshops, organizing literary competitions, etc. With the rise of LLMs and other AI-based tools that can be used within the writing process, our visitors have started asking about the role of AI in literature. Therefore, to fulfill our long-term goal of supporting the creativity of people, we first started offering workshops on utilizing AI tools for creative writing, and, eventually, decided to organize a literary competition focused on human-machine co-creation of short stories.

Competition Settings

Over the past 3 years, we organized three rounds of the competition (see Table 1). The task was to produce a short story with some involvement of some AI tools. The language of the short stories was Czech (with the possibility to create the story in English and translate automatically to Czech). The competition is primarily targeted at Czech and Slovak starting authors, although we do not actively exclude other participants. It is publicized through a wide range of channels, including organizers' websites and newsletters, relevant Facebook groups, mailing lists of participants of writing workshops, other literary competitions, etc. Due to our long-term collaboration with aspiring authors, we are generally well connected with them and can thus reach them rather efficiently. Some participants have taken part in multiple rounds of the competition.

After announcement of each round, the participants have several months to turn in their submissions via e-mail. After that, the submissions are judged by human judges, and authors of the top 5 stories are then invited to a modest ceremony at the library in Prague. The competition has a minimal budget, with a lot of community volunteering, and with symbolic prizes offered by several local sponsors to the top 5 participants in each round (typically books, stationery, and AI-themed artifacts). We are preparing to publish some of the best stories in a dedicated anthology, and they are also currently being published in audio format read by a synthesized voice: `https://talk.youradio.cz/porady/nova-vlna`

The length limit for the stories in the 1st round of the competition was 3600 characters (including spaces), and 5400 characters in the 2nd and 3rd round (i.e. approximately 2-3 pages of text). Each round had a special focus, designed to test various aspects of the co-creation process:

2023: sci-fi, fantasy, horror The first round was designed to investigate the originality of the stories and the ability to meaningfully operate and reason about fictional worlds with different rules of functioning. This is in principle difficult for LLMs, since they tend to generate typical rather than extraordinary, clichés rather than peculiarities (Holtzman et al. 2020); and also since they are typically able to reason only within the rules and functioning of our real world (Wei et al. 2022).

2024: detective story, thriller The second round targeted the ability of LLMs to work with concealed information, known to the creator from the start but only revealed to the reader at the end. This is in principle difficult for LLMs since they have no concealed mechanism to pass on information into future (Salem, Paverd, and Abdelnabi 2026), they can only use the open text (except for modern reasoning models, which were not yet widely available at the time of the 2nd round).

2025: explicit fictional author In the third round, we decided to further explore the paradigm used by the winning submission to the second round: first design an explicit author persona and then condition the generation process on the persona. (The authors were asked to also submit a bio of the fictitious author, of up to additional 1000 characters.) This also plays with the common description/critique of AI-generated texts not having an author (Craig and Kerr 2025) by providing an author with a name and biography, even though non-existent.

As an example, we enclose the winning short story from the 3rd round in the appendix.

Competition Reports

Apart from the co-created stories, we also collected reports in which the participants detailed their co-creative process. In the first round, the participants were instructed to submit a free-form unstructured DOC(X) report specifying the AI tool(s) used, record/description of their interaction with the tools (ideally a full listing of the prompts and interactions), to what extent they feel to be the author of the story (0% to 100%), and a short account of how the interaction with the AI tools was. In the second round, we additionally asked how experienced they are with AI and how experienced they are in fiction writing. In the third round, we also asked to what extent AI was used to create the persona of the author (0% to 100%). We did not collect any other demographic information about the participants.

Analyzing Reports

The reports provide very valuable data for our investigation; however, all the information is self-reported by the participants and thus inherently unreliable to some extent. Unfortunately, there are various gaps, as not all reports contain all the information that we would like to have. Moreover, the reports are not structured in any consistent way. We deliberately decided to make the instructions as simple and low-barrier as possible, and to be rather benevolent with what the participants submit, not to discourage people from participating. This, in turn, leads to the automated analysis of the reports being somewhat complicated.

For this reason, the first phase of the report analysis was done completely manually. One member of the research team manually went through all of the submitted reports,

interpreting the provided information in a structured way and collecting it in a table. In this way, we collected the simpler information that we explicitly solicited (tools used, authorship, writing experience, AI experience), as well as some semi-structured information about the writing process (e.g. structure of collaboration, roles of human and AI in the process, author of the theme, who did the final editing, interesting observations). Categories for frequent responses were designed and updated on the fly as necessary, with still some responses not fitting any of the categories (and some aspects missing altogether in the reports). Subsequently, an LLM pipeline was used to analyze the table and the reports to identify potentially interesting findings that can be based on the data. The main findings suggested by the LLM pipeline were then verified with the data, and further analyses and correlations were computed using standard statistical tools.

All the reported findings are based on our manual or manually verified analyses of the reports and stories (apart from the automated quality evaluation; see later). The strength of our findings and resulting claims is naturally limited, as we are dealing with somewhat noisy and unreliable data processed in a somewhat noisy way (e.g. we needed to interpret the authors' self-reported AI expertise, communicated in various ways, as falling into a limited set of categories, in order to correlate it with the quality of the submissions). In general, we assume that some reports may contain inadvertent inaccuracies or omissions; however, we expect these errors to be random rather than systematic, and therefore unlikely to substantially bias the aggregate analysis. Moreover, we note that the participants had no incentive to lie in their reports, as the reports were not taken into account in the competition itself.

All our data that we are allowed to make publicly available are shared in our GitHub repository at `https://github.com/ufal/aiai`. This includes the calls for participation with author instructions, the top 5 stories from each round, tables with data analyses, and various support scripts. Unfortunately, we are not allowed to publicly share the reports; a shortcoming which we intend to fix with the next competition round, potentially additionally also asking the authors of the previous rounds whether they would consent to making their reports publicly available.

Additionally, as an example of report analysis, we enclose a more detailed analysis of the report for the winning submission from the 3rd round in the appendix.

Quality evaluation by human judges

Within the competition, the stories are independently rated by several human judges (6 judges for the 1st and 3rd round, 7 judges for the 2nd round), with each judge reading all of the submitted stories and ranking the top 21 of them.

Judges were recruited from three areas of expertise (with some judges belonging to multiple of these categories):

Librarians Readers with immense reading experience and vast knowledge of books (but not necessarily any expertise in creative writing).

Authors Both professional and amateur fiction writers (each with at least one published work). They have some degree of writing expertise and thus evaluate not only their impression of the work but also how the author works with the text.

AI experts Experts with understanding of AI functioning, experienced with dealing with LLM outputs.

Mixing various kinds of expertise among the judges ensures that the highest-ranked stories are of a good quality ideally in all relevant aspects.

The judges only have access to the submitted stories but not to the reports. The formulation of the task for the judges is as follows: *'Read the texts as standard short stories and select those that you subjectively like the best'*. Each of the judges independently reads and ranks the stories submitted to the given competition round, and their rankings are then combined into one overall judges score. The top 5 stories are then announced as winners.

The exact mechanism of the ranking is inspired by other Czech literary competitions, using a "pyramid setup": each judge ranks 1 story at rank 1, 2 stories at rank 2, 3 stories at rank 3, etc., up to 6 stories at rank 6. In this way, each judge only ranks the top 21 stories according to their evaluation; this is because it is easier and more important to differentiate between the best texts than between the worst texts.

The ranks are then converted to points using the formula $points = 60/rank$; the assigned ranks 1-6 are thus converted into 60, 30, 20, 15, 12 and 10 points (or 0 for non-ranked stories). The points from individual judges are then summed up for each submission, and the submissions are ordered by the resulting *judges scores*. The nature of the ranking mechanism leads to the fact that there are rarely ties on the top places (which is good for announcing winners), but very often ties at several of the bottom places (which does not matter). In our work, we use the resulting *judges scores* of the stories in each round as the ground truth evaluation of their quality, and correlate it with various other aspects.

Insights from the Judges

Besides the official ranking, many of the judges actively communicated various additional insights and comments with us. Moreover, several of the judges took part in judging multiple rounds of the competition, we can thus gain some absolute as well as relative insights. The judges' comments were communicated in a non-structured informal way; in this section, we summarize interesting insights which we gained through this communication.

In the first round, the quality of the stories was quite low, with a lot of repetitiveness and homogeneity; the second round was somewhat better but not strongly. In the first two rounds, the judges thus agreed that the task of selecting the top 5 stories consisted of identifying stories that were good and/or interesting at least in some aspects, but generally being 'bearable' rather than really good.

However, the third round of the competition showed that the quality of the available AI tools is already high enough to potentially reach the writing quality of a mediocre human writer, even when using simpler approaches. The judges agreed that they observed a significant improvement in quality between the 2nd and 3rd round. All the top 10 stories

were interesting, and the task was to select the 5 best stories among 10 relatively good candidates.

Although all judges knew that the stories are partially generated, in the third round we also had reports that for some of the stories, the reader temporarily forgot that they are reading a generated text. Some judges reported that some of the generated stories elicit thoughts in them, and/or spark emotions; qualities traditionally attributed to human works but assumed unattainable by generated works.

However, there are still qualities that seem hard or impossible to reach with current AI tools without considerable human creative intervention, such as building an intriguing plot with a strong plot twist and/or a sharp point. The LLMs are getting increasingly better at the writing craft, but fail to provide sufficient amount of creative input; the human contribution is still important in this aspect.

Comparison to human literary competitions As many of the judges also have experience in judging competitions of fully human-written texts, they shared some similarities and differences between these and our setting.

First, in both fully-human amateur competitions as well as our human-machine co-generation competition, they observe that the vast majority of the submissions in the competition are dull and uninteresting.

However, AI tools seem to struggle more with aboutness and interestingness than human authors. In the co-generative competition, the good stories are thus those where the human co-author managed to creatively provide these qualities; through various ways though, including providing the creative input at the beginning, and/or selecting sufficiently creative ideas from multiple generated variants, and/or enriching a first dull version of the story with manually provided ideas on making it more interesting and letting the AI tool to iteratively incorporate the human creative input, etc.

On the other hand, in fully human competitions, there is typically a certain amount of submissions where the inexperienced human authors clearly struggle with the craft of writing; this is mostly missing in the co-creation setting, as the current AI tools are masters of the writing craft.

The general sentiment is that the top stories from the 3rd round of the competition could successfully compete with fully human-written texts in standard literary competitions, probably ranking near the top in smaller amateur-focused competitions, and maybe somewhere in the middle in big competitions with participation of professional writers.

Automated Quality Evaluation

We have also run an automated LLM-based evaluation of creativity, to examine usefulness of such automated evaluation in this setting, and to gain more detailed insights into how LLMs perform in various aspects of creative writing and how this changes over the years.

LLM-based evaluation is currently a usual choice for automated evaluation of natural language generation, surpassing previous approaches, although there are also numerous concerns, such as reproducibility of the results (Kartac, Lango, and Dusek 2025). LLMs often demonstrate respectable correlations with human ratings, especially when the evaluation targets semantic properties of the text rather than formalistic properties (Mareček et al. 2025). A common issue may be a high variance and low agreement of the scores, which is often due to too simple prompts; it has been shown that more reliable results can be obtained by prompting the models with explicit checklists of features/qualities that the LLM should explicitly evaluate (Lee et al. 2025).

LLM-based quality	Average '23	Average '24	Average '25	Corr/Score '23	Corr/Score '24	Corr/Score '25
Characteriz.	3.2	4.0	5.7	0.63	0.71	0.29
Plot	3.2	3.9	5.7	0.56	0.69	0.33
Setting	3.3	4.2	6.4	0.57	0.59	0.31
Conflict	3.0	3.9	5.6	0.64	0.67	0.35
Theme	2.9	3.5	5.4	0.52	0.43	0.26
Voice	3.6	4.2	6.2	0.61	0.58	0.36
Craft	3.6	4.2	6.0	0.44	0.66	0.38
Originality	2.5	3.1	5.1	0.65	0.49	0.24
Creativity	2.5	3.1	5.3	0.70	0.48	0.27
Avg score	3.1	3.8	5.7	0.68	0.68	0.34

Table 2: Average automated quality scores per round obtained through LLM prompting (0-10, higher is better) and their Spearman correlation coefficients with judges scores.

Our evaluation prompt is based on *LLM Creative Story-Writing Benchmark V4* (Mazur 2026). The benchmark defines 8 criteria to evaluate the quality of generated stories, evaluated on the range 0-10 (higher is better). Detailed definitions of the categories are used as a prompt for LLMs as judges, with explicit checklists of expected positive properties as well as common failures. The 8 categories evaluated by the benchmark are: Characterization and Motivation; Plot, Structure and Causality; Setting, Atmosphere and Specificity; Conflict and Stakes; Theme and Subtext; Voice and Point of View; Prose and Line-Level Craft; Originality and Ingenuity.

However, none of the categories evaluates creativity as such, although the last category goes in that direction to some extent. Therefore, for our experiments, we enriched the framework with a 9th category specifically targeting evaluation of overall creativity. The creativity dimension of the rubric operationalizes the standard research definition of creativity as the production of artifacts that are both novel and appropriate within a given domain (Runco and Jaeger 2012; Amabile 2018). To capture this construct in narrative texts, the rubric evaluates the presence of original combinations of familiar elements, transformation of genre conventions, productive surprise, and the functional integration of novel ideas into the story's structure and meaning, drawing on theoretical models of creative cognition and creative products (Boden 2004; Simonton 2012; Besemer and O'Quin 1999).

We used *Gemini 3 Pro Preview* LLM. The full prompt (8,253 words) and codes are part of our GitHub repository.

Table 2 lists the results of the automated evaluation; we list the average score of each metric over all stories, average of all the 9 metrics, and the correlation of each of the scores with the judges score (Spearman correlation coefficient).

The correlations are rather strong for the first two rounds, mostly between 0.5 and 0.7 for the individual aspects and nearing 0.7 for the average, showing that the automated evaluation is rather reliable. However, for the last round, the reliability drops dramatically; we assume this is due to the measures being computed with current models, which are good at rating the weaker older models but not for current similarly strong models (Thakur et al. 2025).

Note that the judges did not provide scores stratified along these 9 dimensions; the judges score is one-dimensional, and we thus correlate all of the categories with the same score. As a good story should be good in all of the 9 aspects, it is reasonable to expect a moderate correlation of each of the scores with the judges score, but a very high correlation is not expected as each of the scores only evaluates one of the important aspects. For the same reason, a higher correlation is expected (and is indeed observed) for the average of all the scores. The differences of the individual score correlations between the rounds may also suggest that some of the aspects have different importance in the rounds; e.g. originality and creativity may be more important in the sci-fi/fantasy round than in the detective story round.

We can see that the absolute quality is rather low, with averages between 3-6 points (although individual stories scored up to 8.0). Moreover, the Originality and Creativity scores are among the lowest, confirming that LLMs struggle with the creative aspects of story-writing more than with the more technical skills (the craft). On the other hand, we can clearly see the improvement in quality between the rounds in all aspects.

Human-Machine Interactions

The mode of interaction between the human and the AI tool can have a large impact on the quality of the output as well as the feelings and satisfaction of the human (Issak, Rezwana, and Harteveld 2025). One of our interests is what roles do humans and the AI tools take on in the creative process, how they collaborate, and what are the effects of the approach.

The writing process consists of several steps and corresponding roles (the actual steps may vary for individual authors to some extent). Our general observation here is that both humans and AI tools can take on any of the roles in the creative process; but some scenarios are more typical than others. To achieve high-quality results, human creative input is important *somewhere* in the process; however, there is no single point in the process that would require human contribution. In the competition reports, we have seen a great variety of division of the tasks between the human, the AI tools, or a collaboration of both, and we have not identified any single step in the process that would always be done by the human or always by the machine.

Source of the Story Idea

As an example, we can look in more detail at the ideation role – the invention or emergence of the main idea of the story – which typically takes place somewhere at the beginning of the creative process (but the idea may also emerge and/or evolve later during the process).

Competition year	2023	2024	2025
Initial idea from AI	30%	50%	60%

Table 3: Stories where the initial idea for the story is AI generated, not human-provided; in tens of percent.

In their concept of framework authorship, (Piorecký and Husárová 2024) assume that the human contribution is crucial at the beginning and end of the creative process: they assume that the most important human creative input is in the idea of the story and plot, and that a human is also needed to eventually decide to publish the story.

In the reports, we have encountered various approaches, such as having human creative input at the very beginning of the process, or have the AI tool propose the main idea of the story, or having the AI tool suggest multiple ideas with the human then choosing from them. Our observations thus do not support the assumption of Piorecký and Husárová.

As most of the authors explicitly stated whether the story idea came from the human or the machine, we can even analyze this quantitatively. Table 3 shows the proportion of stories where the idea was AI-generated. We can see that already in the first round, 30% of the submissions were based on an AI-generated idea, with the proportion steadily rising up to 60% for the 3rd round.

The situation is similar with all the other steps and roles in the creation process.

Typical Co-creation Scenarios

Based on the reports, we can approximately describe several generally common approaches, although many submissions deviate to some extent from these prototypical scenarios:

One-click generation The most basic approach is where the human provides only a simple prompt (or even just a copy-paste of the competition task), without much or any creative input, and submits the first output that the LLM generates. This approach is typical for inexperienced AI users and does not often lead to good results.

Human as orchestrator Here, the core of the human creative input is provided at the beginning of the process, defining e.g. the idea, task, style, requirements, preferences, and/or direction. Then, the AI tool is used to convert the detailed instructions and/or ideas into text. Finally, the text is curated and iteratively refined through human intervention by dialogically instructing the AI tool to perform specific improvements. Some participants, especially those with extensive AI experience, use rather complex prompt design and/or combination of several clearly defined steps in the process to carefully guide the generation process step by step (sometimes, even the steps to take are suggested by the AI tool). This is one of the most successful approaches, based on true human-machine collaboration, with a varying degree and kinds of creative contributions of both the human and the AI tool(s).

Human provides detailed input A variation of the human orchestrator scenario, where the focus is not on the complexity of the prompt and setup, but rather the detail of

the provided input. Here, the main creative contribution is by the human, who provides a detailed outline or even a draft version of a text. The AI system is then instructed to convert the input into a well-written story, but the focus is on its mastery of the writing craft rather than its creative input; the AI tool functions as a post-editor and/or ghost-writer rather than a full creative partner. Therefore, the success of this approach is determined by the quality of the provided input, and is thus successful with participants who are themselves able to build a good story plot.

Human as editor Here, the AI tool has the major creative contributions, by being only lightly instructed, having a lot of freedom to decide what to write and how to write it (and typically also writing it). Then, the human takes over as an editor rather than a creative co-author, manually editing the generated text (sometimes even massively). This approach is usual with users with little AI experience but some writing experience. The results are often poor, as the AI-generated draft is often of low quality and thus difficult to turn into a good story. Experience also shows that manually editing the generated text can make it stylistically inconsistent; better results may be achieved by indirect editing through dialogically instructing the AI tool and thus iteratively refining the text.

AI only for auxiliary tasks This is a typical way of using the AI tools by experienced writers. The text is fully written and edited by the human, with AI tools used for auxiliary tasks (brainstorming, research, beta-reading, suggesting improvements, etc.) but not for directly editing the text; the human keeps full control over the text. This approach is thus successful if the participant is already experienced in writing.

Although the approaches are highly varied, we observe that the AI tools are good at the writing craft itself, but usually require human creative input to make the story novel and interesting. We can thus think of the optimal role of the human in the writing collaboration as a sort of creative leadership and curatorship, guiding the creation process but letting the AI system do most of the actual writing.

However, there is, by far, not a single optimal setup; all of the aforementioned approaches were used at least for one of the top 5 stories in at least one round of the competition.

Usefulness of AI tools for creators

Based on the reports, we note that the usefulness of AI tools typically differs for amateur writers and for professionals.

Most of our participants have no or low experience in writing, but many of them express a long-lasting desire to write; they have simply always lacked some of the necessary skills and thus had not been able to write satisfactory texts. For them, the AI tools are enabling, finally making it possible for them to become authors or co-authors of literary texts of reasonable quality, unachievable for them without the support of the tools. This is a sort of prosthetic usage of the tools, compensating for some weaknesses or disabilities on the participant's side.

Experienced writers, who are capable of writing high-quality literature without AI help, do not benefit from in-

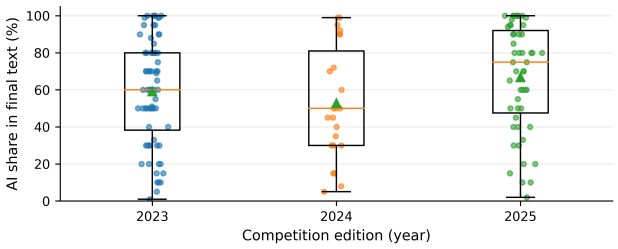

Figure 1: Distribution of the amount of AI contribution.

Competition year	2023	2024	2025
Average AI contribution	59%	53%	67%

Table 4: Average author-reported AI contribution.

volving AI tools directly into the creative process; this rather disrupts the process and leads to worse results. However, they can benefit from AI tools for auxiliary tasks, such as research, brainstorming, proofreading, or critical feedback.

There is an ongoing discussion whether AI tools support or kill creativity in humans (Alegaonkar and Avachat-Shirke 2023; Alzubi, Nazim, and Alyami 2025; Ingle 2025). Based on our observations from the competition, we see that if the AI tools are used adequately, they do not influence the creative process negatively. They can support amateur writers by augmenting them with skills they lack to let them fully utilize their creative potential. In this setting, we do not see the threat of limiting their future creative potential (Kumar et al. 2025), as without the tools, they would not create anything anyway; however, there is always the threat of limiting one's cognitive capabilities in general through reliance on AI tools (Kosmyna et al. 2025), which we do not measure.

For professional writers, it makes little sense to use the tools substitutively to automate skills which they already have, but the tools can nevertheless support their creative process by helping with other related tasks.

Optimal AI/human balance

Looking beyond the actual roles taken on by the humans and the AI tools, we can zoom out to investigate the proportion of AI contribution in the process and its effect on the quality of the results. Bear in mind that the amount of AI contribution is estimated and self-reported by the participants and

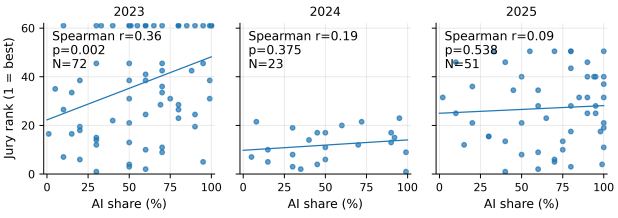

Figure 2: Relation between (self-reported) share of AI on the creation process and final rank in the competition.

Tool	2023		2024		2025	
GPT	57	77%	19	83%	38	72%
Gemini	7	9%	2	9%	8	15%
Claude	0	0%	3	13%	4	8%
Copilot	8	11%	0	0%	3	6%
other	7	9%	2	9%	2	4%

Table 5: Tools used by the participants. Some participants used more than one tool. All versions of the tools are grouped together; most participants used the most advanced version widely available at the time of the competition.

may not be completely trustworthy, all the results in this section thus inherently have only limited reliability.

Figure 1 and Table 4 show the amount of AI contribution over the three rounds of the competition. While the average share of AI rises slightly over time, it varies a lot among the individual submissions and is rather spread out.

Figure 2 analyzes the relation between the share of AI in the process and the final rank of the resulting story in the competition. From this, we can draw some more specific insights, although even here the situation is quite varied.

In general, a meaningful and balanced AI-human collaboration seems to be more successful than extremely large or extremely low AI share; however, the range of optimal AI share lies approximately in a broad range between 30%-80%, and successful outliers still do exist.

In the earlier rounds of the competitions, better results were achieved rather with lower AI contribution, but this effect gets weaker with each round: in 1st round, there was a correlation of 0.36 between low AI share and good rank in the competition, but this drops to 0.19 for 2nd round and 0.09 for third round. It still holds that some human creative contribution is necessary to achieve good results, but the required amount gets smaller each year.

Fully AI-generated stories with reported 0 human contribution all ranked last in the 1st round, whereas in the 3rd round they ranked around 30th place on average. The best fully generated story ranked as high as 11th place, thus already outscoring most stories with some human contribution (it was created by a participant with very high AI experience as well as literary experience).

Tools used

In Table 5, we provide an overview of AI tools used by the participants; the tools grouped under the label 'other' include: you.ai (3x), poe.com, Deeply, WizzardLM, DeepL, Bing, and Perplexity (all 1x). The distribution is rather stable across the rounds, with GPT-based tools being a stable default for most participants; this of course limits our ability to draw any strong findings, as all of the other tools are used only rarely. Nevertheless, we did not observe any significant links between the tool(s) used and the quality of the output; it seems the approach matters more than the tools.

Effect of Experience

We now investigate the effect of the experience of the human along two dimensions: technical experience in using the AI tools, and literary experience in manually writing texts.

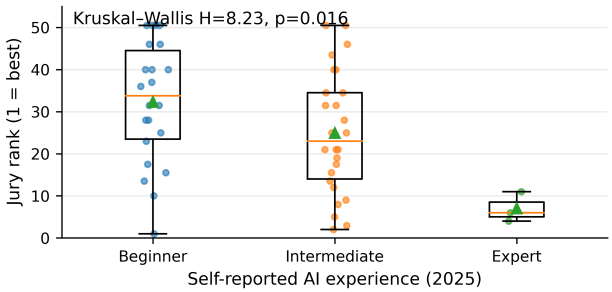

Figure 3: Effect of technical experience

AI experience (self-rep.)	None/ Beginner	Intermediate/ Advanced
All	43%	57%
Top 10	26%	74%

Table 6: Self-reported technological experience, proportion over 2024 and 2025 combined (the proportions were similar in each round). As the detailed categories were slightly different between the years, we combined them into only two coarse-grained categories.

Effect of technical experience and complexity

Our investigations clearly show that experience in using AI tools as well as sophistication of the co-creation process are vital in producing a good output.

Table 6 shows that approximately half of our participants describe themselves as inexperienced in using AI tools, while the other half see themselves as advanced AI users. However, the vast majority of the human co-authors of the top stories are technically experienced. Figure 3 paints an even stronger picture, showing that all of the highly experienced AI users scored very high in the 3rd round (within the first 11 places), while only 2 AI-beginners scored in the top 10. On the other hand, the winning submission of the 3rd round was actually by a self-described AI beginner, showing that the current generation of AI tools can be successfully used even by less experienced users.

In the reports, the participants gradually report that the interaction with the AI tools is getting easier and more pleasant. In the first round, they often reported that using the tools resembled a fight rather than collaboration, requiring a lot of manual work to achieve a bearable result, the process was often frustrating, and the final text was typically still rather disappointing. In the second round, the interaction with the tools got easier, but the results did not improve so much. However, in the third round, the participants were often very satisfied both with the process and with the results. They commonly describe the process as easy, pleasant, with the tools following well their wishes and instructions, and requiring much less manual interventions into the process.

Through manual investigation of the reports, we have also observed that the approaches used by the participants are

Corr. of score with	Approach	AI skills	Lit skills
2023	0.28		
2024	0.42	0.23	0.13
2025	0.06	0.30	0.02

Table 7: Spearman correlation of the judges score of the short story with the complexity of the approach (estimated as the length of the report, in characters), self-reported AI expertise, and self-reported literary writing experience.

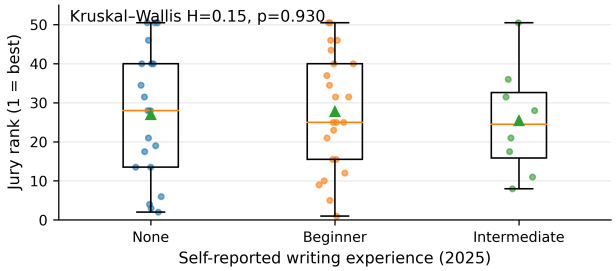

Figure 4: Effect of writing experience

getting more advanced and complex. As an easily quantifiable approximate proxy for approach complexity, we use the length of the reports (number of characters); we observe that the reports get longer by approximately 25% between 2023 and 2024; the length stayed similar between 2024 and 2025.

Table 7 correlates the judges scores of the stories with the approach complexity and technological expertise of the participants, typically showing moderate to strong correlations.

The effect of approach complexity practically disappears in the third round. This is in line with our observation based on the reports that in 2025, even simple one-click approaches are starting to achieve competitive results, suggesting that the AI expertise and approach complexity may become less important in future, as the AI tools are gradually improving and are increasingly successful in following even simple naive instructions.

Effect of writing experience

On the other hand, we did not find any effect of experience with literature writing on the resulting quality of the co-created stories; see the weak to negligible correlations in Table 7 and the very similar distributions of scores among the participants in Figure 4. It thus seems that writing experience is not helpful in co-creation of stories with AI; probably because many of the writing skills are well automated by the LLMs, and it is rather experience with the AI tools which is crucial in the process. Still, we note that the vast majority

Writing exp.	None	Beginner	Experienced
All	39%	47%	14%
Top 10	42%	42%	16%

Table 8: Self-reported writing experience, combined over 2024 and 2025 (the proportions were similar in each round).

Word	Stories containing it	
dark, darkness	45	61%
mystery, mysterious	39	53%
shadow, shadowy	38	51%
night	37	50%
forest	27	36%
whisper	21	28%

Table 9: Number and proportion of stories in the 2023 round, containing the selected genre-typical/cliche words.

of our participants have zero or little writing experience (see Table 8), our data for experienced writers are thus limited.

On closer investigation, we noted that although there is no meaningful difference in the scores of literary amateurs and experts, there are differences in their struggles. What we observed with amateur authors is that if they e.g. do too much editing of a generated text, they often worsen its quality, as they do not have the experience in producing good texts. However, experienced authors struggle differently, as they have too much experience in producing good texts, and thus are not willing to accept a lot of what the AI tools generate. They tend to fight with the AI tools, or even compete with them, unable to accept the AI tool as a creative partner and to meaningfully collaborate with it.

Round-specific Findings

1st Round: struggle with homogenization and fictitious worlds

The first round focused on science-fiction, fantasy and horror stories. We found that the submitted stories typically struggled with originality, to the extent that many of the stories were quite similar to each other, featuring similar plots and other features. Specifically, the most common plot was a horror story of a group of young people finding an abandoned house in a dark forest; upon entering the house, they experience an unpleasant encounter with a mysterious figure (this plot appeared in more than 10 of the stories). We also noted that there are apparently some genre-typical words that appear way too often in the generated stories (see Table 9). The style and structure of the stories was often similar, with a lot of clichés, predictable themes and development, shallow plot, pathos, and a happy ending. All of these are clear examples of the phenomenon of LLM homogenization (Anderson, Shah, and Kreminski 2024), where individual outputs are not necessarily bad on their own, but investigation of a larger amount of generated texts reveals their high similarities in many aspects; a literary competition with a fixed task is exactly a great setting for observing this effect.

In 2023, the AI tools also still struggled with the writing craft in Czech language, with many texts being rather outlines or descriptions of a story than actual engaging stories. We also observed e.g. a low amount of dialogue and other good storytelling practices (we gradually see more and more dialogues embedded in the stories in subsequent rounds).

As for the fictional theme of the round, we observed that LLMs could design fictitious worlds that function dif-

ferently from our world (e.g. by having magic), but then failed to meaningfully utilize the introduced mechanisms (e.g. to use magic to resolve the plot). This is due to the fact that surprisingly, LLMs can work as a simulation of the real world to some extent as they have gained the emergent capabilities of reasoning about our world even without having direct access to it, just based on the texts in the training data talking about the world (Zhang et al. 2025; Ge et al. 2024). However, this is a difficult and complex capability that only emerges with training on a large amount of texts about the world. For the invented fictitious world, the models have not been trained to reason about it – unless it is a fictitious world well-represented in the training data (Shostack 2024) – and are unable to grasp the logic of the just-introduced world only through a zero-shot exposure to its brief description.

2nd Round: Concealed Information

The second round focused on detective stories and thrillers, where we specifically were interested whether the LLMs can work with concealed information, as a typical decoder-based language model has no actively controlled hidden memory and can only explicitly pass information forward through the generated text. Note that this was before reasoning models became widely available (Huang and Chang 2023) which bypass this limitation by generating the 'reasoning' text that does not become part of the returned output and thus can easily work with concealed facts (Cheng et al. 2025).

We found the LLMs to mostly fail, with a tendency to reveal the secret too early. Successful approaches included simulated reasoning with multi-step/hierarchical setups, where the secret was not technically hidden, but it was overt in intermediate generated texts which did not form the final output (e.g. an outline generated as a middle step before generating the story itself) and thus served as a sort of memory for the language model.

The homogenization/cliché problems got weaker compared to the previous round, although still present to some extent (e.g. 4 of the top 5 stories take place in a library).

3rd Round: Explicit Personas

The third round required the participants to work with an explicit persona of the virtual author within the prompt. The approach to creating the persona was as variable as the story generation, with some participants describing the persona fully manually, some using a fully generated description, and some co-creating the description in human-AI collaboration. Nevertheless, conditioning on the explicit persona seemed to lead to more diverse and interesting outputs (Hu and Collier 2024). To some extent, this pushed all of the participants to use a more complex approach, which may explain part of the higher quality of the resulting stories. While the judges' evaluation focused mainly on the stories, there was feedback that even reading the virtual author biographies was rather enjoyable, with some of the biographies being more interesting than the story.

4th round: Humor?

In the 3rd round, some participants managed to successfully incorporate humorous elements into the short stories, confirming the recently emerging findings that LLMs can already produce humorous texts in various languages to some extent (Wu, Weber, and Müller 2025; Inácio and Oliveira 2025; Amir 2025). This was positively noted by some of the judges as a new development, as the previous generation of AI tools was not yet capable of producing satisfactory humorous elements in Czech language. Also, several of the participants suggested to focus on humor in the next round. We thus believe that the time is ripe to explore how humans and machines can co-create humorous texts. Our provisional hypothesis is that current AI tools can already produce humorous elements in the text, but are not yet good for evaluating their success in doing that; we thus believe that the usual role distribution would be utilizing AI creativity to propose humorist elements, followed by employing human capabilities to assess their quality and select the ones that people will actually find both good and funny.

Conclusion

In each of the past three years, we organized a literary competition of short stories co-created by humans and AI tools. In this paper, we analyzed the submitted stories as well as the reports detailing the participants' approaches, investigating observed patterns and correlations.

We found that the progress in capabilities of available AI tools leads to better quality of the co-created stories in each round, with the best submissions currently already reaching at least the level of mediocre human authors. We showed that AI experience of the participant and complexity of the collaborative approach are strongly linked to a high quality of the resulting text, whereas participant's literary experience is irrelevant. However, as the tools are getting more advanced and easy to use, the complexity effect disappears and the technological expertise effect may weaken in future.

We identified and discussed some common collaboration approaches while observing their wide variety, finding that meaningful balanced collaboration between the human and the machine is important. However, any of the creation steps can be automated and there is thus no single point where the human involvement would be required, and the amount of AI contribution versus human contribution to the process can also vary a lot even in the winning stories.

We note that a lot of the input information for our analyses is self-reported by the participants, our resulting claims thus need to be taken with some caution.

Author Contributions

Miloslav Linc was the main organizer of the literary competition and of the manual evaluation of the submissions. Rudolf Rosa was a co-organizer of the literary competition and the main author of the text of the paper. Michaela Liegertová performed the core of the data analysis. All three authors took part in the analysis of the short stories and reports and in the interpretation of our findings.

References

Alegaonkar, A. A., and Avachat-Shirke, M. A. 2023. Is artificial intelligence killing artistic skills in designers. In *International Conference on Emerging Trends in Design & Arts*, volume 4, 109–115.

Alzubi, A. A. F.; Nazim, M.; and Alyami, N. 2025. Do AI-generative tools kill or nurture creativity in EFL teaching and learning? *Education and Information Technologies* 30(11):15147–15184.

Amabile, T. M. 2018. *Creativity in context: Update to the social psychology of creativity*. Routledge.

Amir, O. 2025. Are AI-generated jokes truly original? Charting the "joke space". In *Proceedings of ICCC'25*.

Anderson, B. R.; Shah, J. H.; and Kreminski, M. 2024. Homogenization effects of large language models on human creative ideation. In *Proceedings of the 16th conference on creativity & cognition*, 413–425.

Besemer, S. P., and O'Quin, K. 1999. Confirming the three-factor creative product analysis matrix model in an American sample. *Creativity Research Journal* 12(4):287–296.

Boden, M. A. 2004. *The Creative Mind: Myths and Mechanisms*. London: Routledge.

Chakrabarty, T., and Dhillon, P. S. 2026. Can good writing be generative? Expert-level AI writing emerges through fine-tuning on high quality books. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems*, CHI '26. New York, NY, USA: Association for Computing Machinery.

Cheng, F.; Li, H.; Liu, F.; Van Rooij, R.; Zhang, K.; and Lin, Z. 2025. Empowering LLMs with logical reasoning: a comprehensive survey. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence*, IJCAI '25.

Craig, C., and Kerr, I. 2025. The death of the AI author. In *Robot Law: Volume II*. Edward Elgar Publishing. 250–285.

Ge, Z.; Huang, H.; Zhou, M.; Li, J.; Wang, G.; Tang, S.; and Zhuang, Y. 2024. WorldGPT: Empowering LLM as multimodal world model. In *Proceedings of the 32nd ACM International Conference on Multimedia*, 7346–7355.

Gómez-Rodríguez, C., and Williams, P. 2023. A confederacy of models: a comprehensive evaluation of LLMs on creative writing. In Bouamor, H.; Pino, J.; and Bali, K., eds., *Findings of the Association for Computational Linguistics: EMNLP 2023*, 14504–14528. Singapore: Association for Computational Linguistics.

Holtzman, A.; Buys, J.; Du, L.; Forbes, M.; and Choi, Y. 2020. The curious case of neural text degeneration. In *International Conference on Learning Representations*.

Hu, T., and Collier, N. 2024. Quantifying the persona effect in LLM simulations. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 10289–10307.

Huang, J., and Chang, K. C.-C. 2023. Towards reasoning in large language models: A survey. In *Findings of the association for computational linguistics: ACL 2023*, 1049–1065.

Inácio, M. L., and Oliveira, H. G. 2025. A full pipeline for context-aware pun generation. In *Proceedings of ICCC'25*.

Ingle, R. V. 2025. An AI tool: Savior or destroyer of human creativity in art and literature. *Unlocking Potential: The Intersection of Human Creativity and AI* 59.

Ismayilzada, M.; Stevenson, C.; and van der Plas, L. 2025. Evaluating creative short story generation in humans and large language models. In *Proceedings of ICCC'25*.

Issak, A.; Rezwana, J.; and Harteveld, C. 2025. MOSAAIC: Managing optimization towards shared autonomy, authority, and initiative in co-creation. In *Proceedings of ICCC'25*.

Kartac, I.; Lango, M.; and Dusek, O. 2025. OpeNLGauge: An explainable metric for NLG evaluation with open-weights LLMs. In Flek, L.; Narayan, S.; Phuong, L. H.; and Pei, J., eds., *Proceedings of the 18th International Natural Language Generation Conference*, 292–337. Hanoi, Vietnam: Association for Computational Linguistics.

Kosmyna, N.; Hauptmann, E.; Yuan, Y. T.; Situ, J.; Liao, X.-H.; Beresnitzky, A. V.; Braunstein, I.; and Maes, P. 2025. Your brain on ChatGPT: Accumulation of cognitive debt when using an AI assistant for essay writing task.

Kumar, H.; Vincentius, J.; Jordan, E.; and Anderson, A. 2025. Human creativity in the age of LLMs: Randomized experiments on divergent and convergent thinking. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, CHI '25. New York, NY, USA: Association for Computing Machinery.

Lee, Y.; Kim, J.; Kim, J.; Cho, H.; Kang, J.; Kang, P.; and Kim, N. 2025. CheckEval: A reliable LLM-as-a-judge framework for evaluating text generation using checklists. In Christodoulopoulos, C.; Chakraborty, T.; Rose, C.; and Peng, V., eds., *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, 15771–15798. Suzhou, China: Association for Computational Linguistics.

Mareček, D.; Hodková, K. M.; Musil, T.; and Rosa, R. 2025. Evaluation of generated poetry. In Akter, M.; Chowdhury, T.; Eger, S.; Leiter, C.; Opitz, J.; and Çano, E., eds., *Proceedings of the 5th Workshop on Evaluation and Comparison of NLP Systems*, 109–118. Mumbai, India: Association for Computational Linguistics.

Mazur, L. 2026. LLM creative story-writing benchmark. `https://github.com/lechmazur/writing`.

Piorecký, K., and Husárová, Z. 2024. *The culture of neural networks: Synthetic literature and art in (not only) the Czech and Slovak context*. Nakladatelství Karolinum.

Porter, B., and Machery, E. 2024. AI-generated poetry is indistinguishable from human-written poetry and is rated more favorably. *Scientific Reports* 14(1):26133.

Runco, M. A., and Jaeger, G. J. 2012. The standard definition of creativity. *Creativity Research Journal* 24(1):92–96.

Salem, A.; Paverd, A.; and Abdelnabi, S. 2026. Stateless yet not forgetful: Implicit memory as a hidden channel in LLMs. *arXiv preprint arXiv:2602.08563*.

Shostack, A. 2024. The boy who survived: Removing Harry

Potter from an LLM is harder than reported. *arXiv preprint arXiv:2403.12082*.

Simonton, D. K. 2012. Taking the U.S. patent office criteria seriously: A quantitative three-criterion creativity definition and its implications. *Creativity Research Journal* 24(2–3):97–106.

Teleki, M.; Bengali, V.; Dong, X.; Janjur, S. T.; Liu, H.; Liu, T.; Wang, C.; Liu, T.; Zhang, Y.; Shipman, F.; et al. 2025. A survey on LLMs for story generation. In *Findings of the Association for Computational Linguistics: EMNLP 2025*, 13954–13966.

Thakur, A. S.; Choudhary, K.; Ramayapally, V. S.; Vaidyanathan, S.; and Hupkes, D. 2025. Judging the judges: Evaluating alignment and vulnerabilities in LLMs-as-judges. In Arviv, O.; Clinciu, M.; Dhole, K.; Dror, R.; Gehrmann, S.; Habba, E.; Itzhak, I.; Mille, S.; Perlitz, Y.; Santus, E.; Sedoc, J.; Shmueli Scheuer, M.; Stanovsky, G.; and Tafjord, O., eds., *Proceedings of the Fourth Workshop on Generation, Evaluation and Metrics (GEM²)*, 404–430. Vienna, Austria and virtual meeting: Association for Computational Linguistics.

Wei, J.; Tay, Y.; Bommasani, R.; Raffel, C.; Zoph, B.; Borgeaud, S.; Yogatama, D.; Bosma, M.; Zhou, D.; Metzler, D.; et al. 2022. Emergent abilities of large language models. *arXiv preprint arXiv:2206.07682*.

Wu, Z.; Weber, T.; and Müller, F. 2025. One does not simply meme alone: Evaluating co-creativity between LLMs and humans in the generation of humor. In *Proceedings of the 30th International Conference on Intelligent User Interfaces*, 1082–1092.

Zhang, X.; Li, S.; Hauer, B.; Shi, N.; and Kondrak, G. 2023. Don't trust ChatGPT when your question is not in English: A study of multilingual abilities and types of LLMs. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, 7915–7927.

Zhang, X.; Lin, J.; Mou, X.; Yang, S.; Liu, X.; Sun, L.; Lyu, H.; Yang, Y.; Qi, W.; Chen, Y.; et al. 2025. Socioverse: A world model for social simulation powered by LLM agents and a pool of 10 million real-world users. *arXiv preprint arXiv:2504.10157*.

Zhao, Z.; Song, S.; Duah, B.; Macbeth, J.; Carter, S.; Van, M. P.; Bravo, N. S.; Klenk, M.; Sick, K.; and Filipowicz, A. L. 2023. More human than human: LLM-generated narratives outperform human-LLM interleaved narratives. In *Proceedings of the 15th Conference on Creativity and Cognition*, 368–370.

Appendix: Example of a short story

Example of a short story from the competition; specifically, the winning submission from the 3rd round, *Květiny se nevrací* (*Flowers Never Return*), by Markéta Fryaufová.

First, we provide a more detailed analysis of the author's report of their creative process.

Then, the full story is provided first in its original form in Czech language, and then as an automated translation by ChatGPT-5.5. The text consists of the bio of the fictional author, and then the story itself. The author (the human co-author) agreed with the text of the story to be made publicly available.

Analysis of the author's report

The author listed the AI contribution at 40% in writing the story and at 15% in designing the fictional author: "Designing the author is mostly my work. (...) The main theme and narrative development of the story, the emotions, and the specific wording in most of the letters come from my own mind. AI helped with the composition and rhythm. It further developed the inserts and letters, which I then revised."

The author reports having some experience with AI, but not for creative writing. She has written some short stories privately for herself but none were published.

The author also shared the full conversation with ChatGPT: `https://chatgpt.com/share/688b380b-aa10-8010-980a-9e9b7c96e9c2`

The conversation reveals various details behind the story that are never explicitly said in the story but only indirectly hinted at. The conversation also shows that the process is genuine collaboration, with many interactions between the user and the AI tool (70 request-response pairs), inspiring each other and gradually forming and refining the text.

The interaction mainly falls into the "human as editor" type, with most of the text generated in its first version by the tool, then manually edited by the human, and finally collaboratively refined. The initial input consists mainly of a rather detailed vision of the fictional author, and a general setting for the story; the storyline is then developed collaboratively, typically by the AI tool offering some suggestions, and the human selecting from the suggestions and refining them according to their preferences.

ChatGPT is also used to provide suggestions and options on how to proceed, to ask about what to do next, and to provide feedback on the text. The human provides a lot of feedback about style, atmosphere, setting, etc., pushing the AI tool to make the story more approachable, believable, easy to identify with, as well as to shape it according to the visions of the human about what it should be like. At the end, the AI tool is used to suggest how to shorten the story to fit within the given length limits, but the final edit is produced by the human.

Interestingly, the main plot twist of the story is created collaboratively, with ChatGPT inventing several options for a plot twist to explain a seemingly non-sensical aspect of the story, and the user selecting one of the options but modifying it to make it even more dramatic.

Růžena (nar. 1946)

Píše o ženách, co zůstaly samy s dětmi v bytech a v tichu. Její příběhy jsou prosté, ale o to silnější. Místo okázalých gest popisuje drobnosti každodenních chvil. Popel v cukřence, studený radiátor a smích co zaznívал do prázdna. Nyní, když její roky hlásí dávno po čtyřicítce objevuje své dopisy z mládí. Růža svými texty připomíná, že paměť má tvar prázdného hrníčku na stole a že největší dramata se často dějí v kuchyni mezi dvěma nádechy.

Vyprávi, aby nezapomněla.

Prolog

Někdo za ní zavřel dveře, aniž by řekl sbohem.

Na perónu to páchlo po dešti a uhlí. Kufr se jí zarýval do ruky, v druhé držela tašku s buchtama. Malý spal, hlavičku opřenou o její rameno, a ona si v duchu opakovala, že to zvládne.

„Neotáčej se," řekl táta, ale ona se stejně otočila.

Okno v patře bylo pootevřené. Věděla, že za ním už nikdo nestojí.

Strojvedoucí zapískal.

Růža si sedla k oknu a položila hlavu na sklo. Pršelo. Z kapsy jí čouhal kapesník, co jí ráno matka vrazila do ruky, když naposled zabouchla vrátka.

I.

Milá mami,

jsme tu.

Byt máme dvoupokojový a jsme v přízemí. Smrdí tu to novotou a lepidlem, i když už tu někdo před námi bydlel. Záchod máme sice na chodbě, ale aspoň vlastní. Zatím tu znám akorát jednu paní ze třetího patra. Říkají jí Irenka, ale na Irenu nevypadá. Hulí jak tovární komín a hází vajgly z okna. Pes jí chodí kadit ke sklepům, prý je to normální. Venku je všechno šedý a vzduch páchne po síře. Dole pod kopcem hučí továrna, asi mně brzo pukne hlava. Já nevím. Zatím spíme špatně, protože se Malý pořád budí, asi to tu nepoznává. Ptá se, kdy pojedeme domů... Co mám odpovědět? Když jsme přijeli, měla jsem pořád hlínu za nehty a v tašce buchty od tebe. Všechno to najednou vypadalo cizí. Ale ty buchty jsem ohřála. A víš co? Honza pořád voní po seně. Taky už je zapsanej u pluku a zítra jde do služby. Musím si najít práci a Malého dát do jeslí, promiň, vím že bys nesouhlasila. Ale tu práci potřebuju, abychom to do začátku utáhli. Napiš, co děláte. A jestli vám chybím aspoň trochu.

Tvoje Růža

Sedím u okna, vítr mlátí do parapetu a někde pod námi řve rádio. U nás se pořád řve. Látkové záclony se nadzvedávají ve větru a já potahuji z cigarety. Kouř stoupá z bytu ven, stoupá výš, až nad parapet do noci. Voní vyprané prádlo, jako tehdy, když jsme sem poprvé přijeli. Jen ten pach lepidla a betonu je pryč, teď ho nahradila zatuchlost záclon a cigaretový dým. Měla bych se svléknout, říkám si, vyprat šaty, odnést flašky a vyměnit ty podělaný záclony. Místo toho ale zašmátrám pod stolem a vytáhnu obálky ze zažloutlého papíru. Prsty přejíždím po stránce, ale zdá se mi tenčí, než bývala. Okraj je navlhlý a levý růžek se jaksi

ohnul, asi jak jsem je kdysi neomaleně skládala do kredence. Potáhnu po druhé. Řádky se míhají jako titulky starýho filmu bez zvuku. Z protějšího bytu doléhají zvuky večerních zpráv.

Co bys na to asi řekla?

II.

Milá mami,

píšu ti mezi praním a polívkou. Prací prášek tu smrdí jinak než doma. Mám z toho zedřený ruce, ale aspoň je tu teplo – teda když funguje topení. Radiátory si tu dělaj, co chcou. Ráno je zima jako v kostele a večer zas hic jak v lázních. Když šel Malý poprvé do jeslí, tak mi brečel do bundy, že chce k babičce. Já vím, že bys řekla, ať si ho nechám doma. Ale kdo nás uživí? Honza je pořád pryč – prý noční, školení, prý že mu to nemám brát. Ale co si z toho vzít, když přijde domů a cítím z něj voňavku, co doma nemám? Našla jsem práci v prádelně. Je to hned u trati, rachotí to tam jak peklo. Dvanáctky za pásem, ale baby tam jsou hodný. Jedna, co tam dělá už třicet let, mi říkala, že „tohle město ženský semele, když nebudeš držet hubu a záda rovně." Tak zatím držím. A záda mě teda bolí slušně. Víš, co je zvláštní? Když večer věším prádlo, nejvíc mi voní tvoje utěrka. Měla jsem ji v krabici s hrnkama a načichla domovem. Možná blbost, ale když ji držím, dýchá se mi líp.

Napiš, co dělá táta. A jestli máš pořád ty fialový astry u zápraží. Taky napiš, jak se má Hana a jestli jí to s tím Jirkou konečně klape. Myslíš, že se budu moct brzo podívat domů? Aspoň na otočku. Pozdravuj všechny.

Tvoje Růža

Stůl trochu vrže, když na něj dosedám. Cigareta mi hoří mezi prsty a popílek padá na lino, mezi drobky od večeře, kterou jsem ani neuvařila. Jednu nohu mám pod sebou, druhou nataženou, opřenou o studený radiátor, co nikdy netopí, když má. V pokoji je ticho. Jen v dálce duní vlak. Ten noční, co jezdíval kolem prádelny. Vždycky jsem si myslela, že do něj jednou nasednu a pojedu pryč. Na stole přede mnou leží dopis. Je rozepsaný, se skvrnou od polívky nebo od kávy. Když ho beru do ruky, trochu se mi chvějí prsty. Byla jsem tak unavená. Všechno se tenkrát drolilo. Malý brečel a Honza voněl cizími parfémy. A tak jsem ti psala, jako bych tím snad mohla všechno vrátit zpět.

III.

Milá mami,

Malý má teplotu a pořád pláče. V prádelně mi dali neplacený volno, prý ať si to doma srovnám. Tak jsem umyla okna, vytřela sklep a vyměnila záclony. Teď je večer a ticho, jen slyším ten vlak, co vyjíždí v osm, protože poskakuje po kolejích jak splašený. Honza se doma moc nezdržuje. Řekla bych ti víc, ale nač. Víš, jak to je, ženská to pozná. Malý dnes nakreslil náš dům. Jenže místo dveří nakreslil plot. „Tudy se nechodí," řekl mi, a tak jsem to přelepila sluníčkem. Někdy si říkám, jestli jsem to neuspěchala. Ale pozdě bycha honit. Pořád mám tu tvoji utěrku. Dneska na ní usnula kočka z vedlejšího bytu. Možná taky ví, odkud voní klid. Napiš, jak je. A jak se má Hanka s malejma. Taky napiš, jestli už začalo pršet. Tady je sucho.

Tvoje Růža

Dnes v noci byla bouřka. Blesky řezaly z nebe jako tenkrát v létě, kdy nespadlo kapek a pak najednou-prásk. S Malým jsme věšeli prádlo na balkoně a najednou museli utíkat dovnitř. Teď už tu ale není, prádlo ano, ale smích dětské tváře nikoliv. Sedím pořád u okna. Cigarety mám schované v cukřence, kam jsem dřív dávala lentilky. Jednu si beru, jako by to byl rituál. Zapálím, potáhnu a snažím se přehlušit ticho, které sedá na nábytek jako prach. Ten třetí dopis… Ten jsem psala v noci. Malý stonal a Honza byl pryč. Nevěděla jsem, co dřív. Vzpomínám si, jak jsem utírala z podlahy rozlitý čaj, a přitom si šeptala věty, co bych ti chtěla napsat. Jenže si je nikdy nečetla. Nikdo je nečetl. Jen já. Pořád dokola.

IV.

Milá mami,

dneska jsem si koupila nový záclony. Bílý, s takovýma malýma kytičkama, co vypadají jako ty astry, co jsi sázela podél plotu. Pověsím je hned, jak se ochladí. Honza už se neukazuje. Prej že: „takhle se žít nedá", no bodejť. Sbalil si igelitku a pak ani neřekl Malýmu sbohem. Ale upřímně, možná je to tak lepší. Co by tu s ním? Sám měl furt jen noční. Malý… vlastně už ne Malý. Psali jsme si naposled před dvěma měsíci. Prý má práci ve skladu. Odjel přes hranice, mami. A zůstal tam. Ani se neohlídl. Ale víš co? Já mu to nezazlívám. Ještě má šanci.Tady už se jen čeká. Na jaro. Na vlak. Na něčí kroky za dveřma, který stejně nikdy nezazní. Občas přemýšlím, jestli jsi to nakonec neměla lehčí než já. Nemusela jsi vidět, co se se mnou stalo. Jak člověk vysychá a praská, když ho nikdo nezalije. Někdy mám chuť otevřít okno a vylít ven všechny ty roky. Jenže bych musela začít sama sebou. A na to ještě nemám.

Však víš, o čem mluvím, že ano, matko

S láskou,

Tvá Růža

Záclona se pohne, i když vítr dávno ustal. Možná jen moje noha, co visí z parapetu, zavadila o látku, když jsem zhasínala cigaretu o kraj popelníku. Prsty mám zažloutlé. Prsty mám staré. Jako by celý život byl jen tenkou vrstvou popela, která se na mě za ty roky nalepila. Dívám se do tmy, ale nevidím nic. Ani sebe. Ani tebe.

Dopisy teď leží rozložené po stole, všechny čtyři. Slova, co se v nich lámou, tekla z ruky, která ještě věřila. Psala jsem ti je, protože jsem měla pocit, že mě podržíš, že mi odpovíš, že třeba i jen písmenem naznačíš, že jsem to nezvorala úplně. Ale tys už byla pod hlínou. Zemřelas dva dny před tím, než jsme odjeli. Táta mi to řekl na peróně. Nechtěl, abych si to rozmyslela. Nechtěl, abych zůstala. A já tedy jela, s kufrem a ztuhlou čelistí. A pak jsem ti psala ty dopisy. Jako bys byla jen moc daleko, jako by existovala pošta až do nebe.

Típnu cigaretu a zvednu se. Sáhnu po záclonách.

A konečně – je strhnu.

Navždy Tvá, Růža

Růžena (b. 1946)

She writes about women left alone with children in apartments and silence. Her stories are simple, and all the more powerful because of it. Instead of grand gestures, she describes the tiny details of everyday moments. Ashes in the sugar bowl, a cold radiator, and laughter echoing into emptiness. Now, long past her forties, she discovers letters from her youth. Through her texts, Růža reminds us that memory has the shape of an empty mug on a table, and that the greatest dramas often unfold in the kitchen between two breaths.

She tells stories so she will not forget.

Prologue

Someone closed the door behind her without saying goodbye.

The platform smelled of rain and coal. The suitcase dug into her hand, and in the other she carried a bag of sweet buns. The little one was asleep, his head resting on her shoulder, and she kept repeating to herself that she would manage.

"Don't turn around," her father said, but she turned anyway.

The upstairs window was slightly open. She knew no one was standing behind it anymore.

The train driver blew the whistle.

Růža sat by the window and rested her head against the glass. It was raining. A handkerchief stuck out of her pocket, the one her mother had pressed into her hand that morning when she closed the garden gate for the last time.

I.

Dear Mom,

we are here.

The apartment has two rooms and we are on the ground floor. It smells of newness and glue, even though someone lived here before us. The toilet is out in the hallway, but at least it is ours. So far I only know one woman from the third floor. They call her Irenka, though she doesn't look like an Irena. She smokes like a factory chimney and throws cigarette butts out the window. Her dog goes to shit by the cellar doors, apparently that is normal here. Everything outside is gray and the air reeks of sulfur. Down below the hill the factory roars, I think my head will burst soon. I don't know. We sleep badly so far because the little one keeps waking up, I think he does not recognize this place. He asks when we are going home... What am I supposed to answer? When we arrived, I still had dirt under my nails and your sweet buns in my bag. Suddenly everything looked foreign. But I warmed up the buns. And you know what? Honza still smells of hay. He has already registered with the regiment and tomorrow he starts duty. I have to find work and put the little one into daycare, forgive me, I know you would not agree. But I need the job so we can survive these first months. Write and tell me what you are all doing. And whether you miss me at least a little.

Your Růža

I sit by the window, wind beating against the sill, and somewhere below us a radio is blaring. Everything here is always blaring. The fabric curtains lift in the wind while I drag on a cigarette. Smoke rises out of the apartment, higher and higher above the sill into the night. It smells of freshly washed laundry, just like when we first arrived here. Only the smell of glue and concrete is gone now, replaced by stale curtains and cigarette smoke. I should undress, I tell myself, wash the clothes, take out the bottles, and replace those damned curtains. But instead I grope beneath the table and pull out the yellowed envelopes. My fingers slide over the page, though it feels thinner than it used to. One edge is damp and the left corner bent somehow, probably from the way I shoved them carelessly into the cabinet years ago. I take another drag. The lines flicker like subtitles from an old silent movie. Evening news drift in from the apartment across the hall.

What would you say to all this, I wonder?

II.

Dear Mom,

I am writing between doing the laundry and making soup. The washing powder smells different here than at home. My hands are rubbed raw from it, but at least it is warm here, well, when the heating works. The radiators do whatever they want. In the morning it is cold as a church and by evening hot as a spa. When the little one went to daycare for the first time, he cried into my coat saying he wanted his grandma. I know you would say I should keep him at home. But who would feed us? Honza is always away, night shifts, training, he says I should not take it personally. But what am I supposed to think when he comes home smelling of perfume we do not have here? I found work in a laundry plant. It is right by the tracks, the noise there is like hell itself. Twelve hour shifts at the line, but the women there are kind. One of them, who has worked there for thirty years, told me: "This city grinds women down if you do not keep your mouth shut and your back straight." So for now I am holding on. And my back hurts properly already. You know what is strange? When I hang laundry in the evening, your kitchen towel smells the nicest of all. I had it packed in a box with mugs and it soaked up the scent of home. Maybe it is silly, but when I hold it, I can breathe easier.

Write and tell me what Dad is doing. And whether you still have those purple asters by the doorstep. Also tell me how Hana is and whether things are finally working out with Jirka. Do you think I might be able to come home soon? Even just for a quick visit. Send my regards to everyone.

Your Růža

The table creaks a little when I sit down at it. A cigarette burns between my fingers and ash falls onto the linoleum among the crumbs from a dinner I never even cooked. One leg tucked beneath me, the other stretched out against the cold radiator that never heats when it should. The room is silent. Only the distant rumble of a train. The night train that used to pass by the laundry plant. I always thought that one day I would board it and leave. The letter lies on the table before me. Half written, stained with soup or coffee. When I pick it up, my fingers tremble a little. I was so tired back

then. Everything was falling apart. The little one cried and Honza smelled of other women's perfumes. So I wrote to you, as if somehow it could bring everything back.

III.

Dear Mom,

the little one has a fever and keeps crying. The laundry plant gave me unpaid leave, they said I should sort things out at home. So I washed the windows, mopped the cellar, and changed the curtains. Now it is evening and quiet, except for that train leaving at eight because it rattles along the tracks like something possessed. Honza hardly stays home anymore. I could tell you more, but what would be the point. You know how it is, a woman can tell. Today the little one drew our house. But instead of a door he drew a fence. "You do not go through here," he told me, so I covered it up with a drawing of the sun. Sometimes I wonder whether I rushed into all this too quickly. But it is no use crying over spilled milk. I still have your kitchen towel. Today the neighbor's cat fell asleep on it. Maybe it knows where the smell of peace comes from too. Write and tell me how you are. And how Hanka and the little ones are doing. Also tell me whether it has started raining there yet. Here it is dry.

Your Růža

There was a storm tonight. Lightning sliced across the sky like that summer long ago when not a single drop fell and then suddenly, crack. The little one and I were hanging laundry on the balcony and suddenly had to run inside. But he is no longer here now. The laundry is, but not the laughter of a child's face. I still sit by the window. I keep my cigarettes hidden in the sugar bowl where I once stored chocolate candies. I take one as if it were a ritual. I light it, inhale, and try to drown out the silence settling onto the furniture like dust. That third letter... I wrote it at night. The little one was sick and Honza was gone. I did not know what to do first. I remember wiping spilled tea off the floor while whispering sentences I wanted to write to you. But you never read them. No one did. Only me. Again and again.

IV.

Dear Mom,

today I bought new curtains. White ones with tiny little flowers that look like the asters you planted along the fence. I will hang them as soon as the weather cools down. Honza no longer shows up. He said, "You cannot live like this," well no kidding. He packed a plastic bag and did not even say goodbye to the little one. But honestly, maybe it is better this way. What would he do here anyway? He only ever had night shifts. The little one... actually not so little anymore. We last wrote to each other two months ago. He says he has work in a warehouse. He crossed the border, Mom. And he stayed there. He did not even look back. But you know what? I do not resent him for it. He still has a chance. Here all people do is wait. For spring. For the train. For footsteps behind the door that never come. Sometimes I wonder whether in the end you had it easier than I did. You did not have to see what became of me. How a person dries out and cracks when no one waters them. Sometimes I want to open the window and pour all those years out into the street. But I would have to begin with myself. And I still cannot do that.

You know what I mean, don't you, Mother

With love,

Your Růža

The curtain moves even though the wind stopped long ago. Maybe it was only my foot hanging off the sill brushing the fabric as I stubbed my cigarette out against the edge of the ashtray. My fingers are yellowed. My fingers are old. As though my whole life were only a thin layer of ash that has stuck to me over the years. I stare into the darkness, but I see nothing. Not myself. Not you.

Now the letters lie spread out across the table, all four of them. Words breaking apart inside them, flowing from the hand of someone who still believed. I wrote them to you because I thought you would support me, that you would answer, that perhaps with even a single line you would hint that I had not ruined everything completely. But you were already under the ground. You died two days before we left. Dad told me at the station. He did not want me to change my mind. He did not want me to stay. So I left, with a suitcase and a stiff jaw. And then I wrote those letters to you. As if you were only very far away, as if there were a postal service reaching all the way to heaven.

I stub out the cigarette and stand up. I reach for the curtains.

And finally, I tear them down.

Forever Yours, Růža