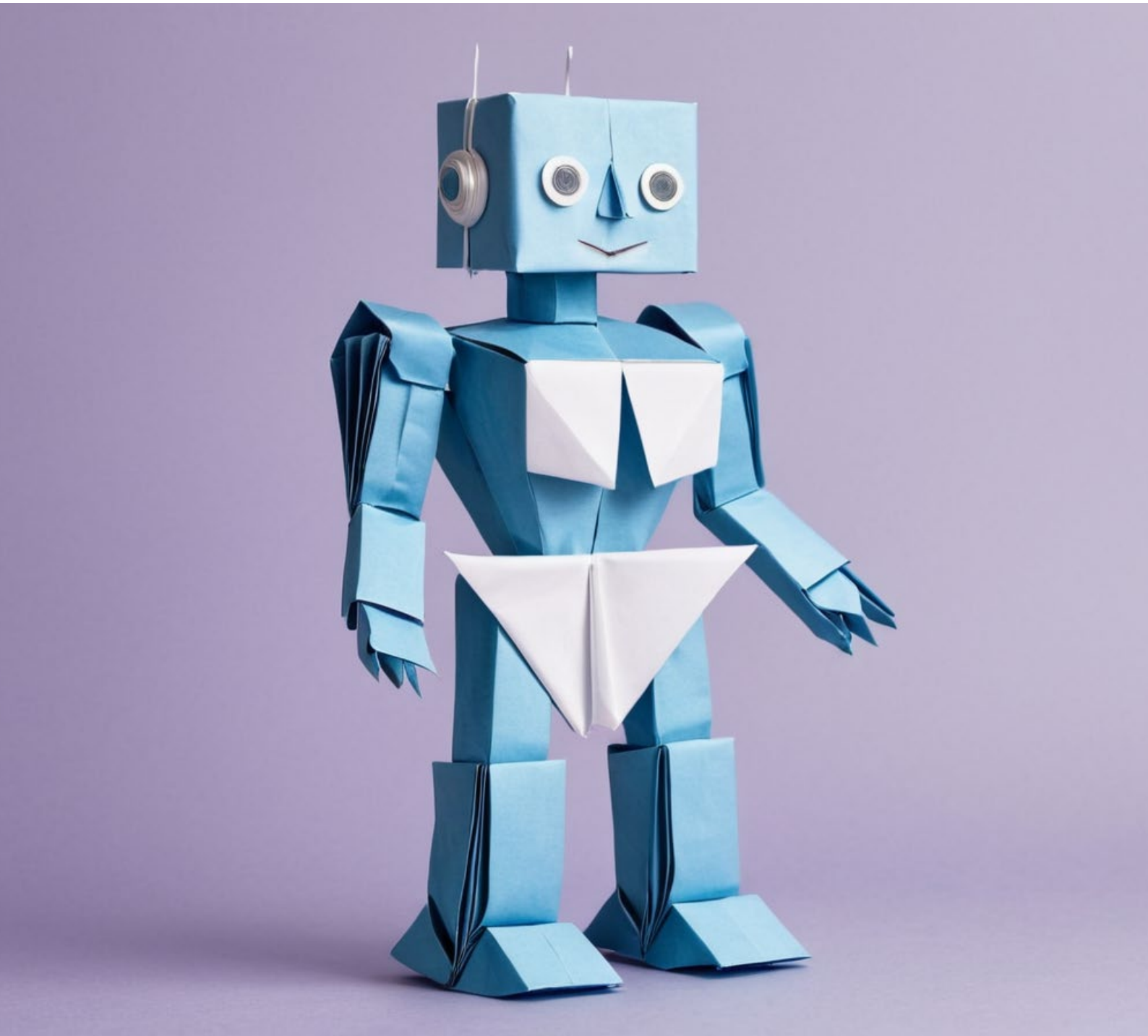
FZI

— Generative AI and its use in a professional environment



– Content

1 Introduction

The rapid development of Artificial Intelligence (AI) has, in recent years, given rise to an innovative class of systems known as Generative AI (GAI). This novel technology demonstrates the potential not only to optimize existing ways of working but also to fundamentally transform the foundations of our interaction with computers and machines. Generative AI goes beyond the traditional application of artificial intelligence by being capable of independently creating content and information without being limited to predefined patterns or datasets. This revolutionary characteristic opens up a wealth of opportunities in professional contexts, ranging from the automation of repetitive tasks to the creative generation of content.

Why is Generative AI so revolutionary?

Generative AI marks a turning point in the history of artificial intelligence, as it possesses the ability to think creatively and generate novel solutions. Unlike conventional AI systems, which rely on training data and pattern recognition, GAI is capable of producing content that goes beyond what has been learned. This autonomy not only allows for improved adaptation to specific professional requirements but also fosters the emergence of innovations that would be difficult to achieve through traditional methods.

Another decisive aspect of this revolution lies in the flexibility of Generative AI. These systems are capable of covering a wide range of tasks and scenarios, enabling diverse applications in professional contexts. From text creation and graphic design to software code development—Generative AI is revolutionizing the way we understand and perform work.

This white paper will explore the fundamental importance of Generative AI in professional contexts, highlighting the various areas of application, challenges, and potential benefits of this groundbreaking technology.

Anyone with some experience with texts written by a generative AI (GenAI) will already have guessed: the 'author' of this introduction is not a human being but a generative AI. In the present case, ChatGPT[1] has explained itself in the original German text and shown the specific properties of a generative AI. All it needed was a short instruction:

>_*I would like to write a white paper on the topic of generative AI and its use in professional contexts. Please write a concise introduction to the topic of generative AI and explain why the subject is so revolutionary.*

The introduction from the German text, as well as the original prompt, has been translated into English by ChatGPT.

Even though GenAI is now producing impressive results, these tools should be used with care, especially professionally. The simple use and surprisingly good results can, at first sight, easily hide that the use of GenAI is linked to problems and pitfalls.

This white paper will consider the potential of GenAI and its challenges:

- legal aspects
- major restrictions
- clarification of functional requirements

[1] https://chatgpt.com

- technical implementation and efficiency

The white paper intends to give you a short overview of the complex subject of GenAI and show you the challenges. Contact us if you plan to use GenAI for your organization's products and processes. We will be happy to consult and support you in profitably integrating GenAI.

2 Legal aspects

The rapid development of GenAI does not leave the law untouched. The following chapter will explain the challenges and changes emerging in the conflict between artificial intelligence and legislation. This section will be adapted continuously to the changing legal framework.

2.1 Copyright

Copyright refers to a work's creator, not to the work itself (principle of creativity). According to § 7 of the German copyright law, the creator of a (piece of) work owns its copyright and is therefore its originator. Copyright protects the creator, the intellectual and personal relation to the work, and the creator's commercial interests. A fundamental condition for copyright is classifying a work as a personal intellectual creation (§ 2, subparagraph 2, German copyright law). Only then is the work protected by copyright.

2.1.1 Training of an AI

The problem in training an AI is not that the AI can 'see' copyrighted work, but in storing the training data and the following duplication of this work. This infringes the exclusive reproduction right of the originator (§§ 15, subparagraph 1.1, § 16 German copyright law).

One way to use copyrighted material for training is to purchase a license. However, this can be challenging with large amounts of data. A further option is offered by the restriction in § 44b of the German copyright law that regulates data and text mining. Text and data mining means the 'automated analysis of individual or multiple digital or digitized works to obtain information, particularly about patterns, trends and correlations' (§ 44 subparagraph 1 German copyright law).

The condition for data mining is that the creator does not contradict its use explicitly in machine-readable form and that the deletion rules are observed. However, labeling online accessible works with a reservation of use by the copyright owner can be difficult. It is therefore recommended that potential training data be collected with particular care.

2.1.2 Prompts

These inputs and instructions to GenAI can trespass the hurdle of copyright law, as can general texts. For example, this can be the case if the formulated prompts go beyond mere work instructions and are characterized by a particular degree of individuality and creativity. However, this does not result in a derived copyright of the product generated by GenAI.

It should also be noted that work instructions for GenAI can be processed using self-authored texts and third-party images or texts. In this case, it must be checked whether one has the necessary rights of use.

2.1.3 Output

Currently, AI-generated products do not enjoy copyright protection. Copyright, based on fundamental rights, is a protective right for human and intellectual creation (Wandtke/Bullinger, § 7 Rn. 18). In principle, the products of GenAI are, therefore, in the public domain.

AI's imitation of a particular style is not legally relevant (Wandtke/Bullinger, Copyright law, § 2 Rn. 40). However, it can be problematic if the output of a GenAI contains specific copyrighted works, as there could be a copyright-relevant duplication or editing. The European Court of Justice assessed the use of a copyright-protected audio track by examining the recognizability of the specific work. It examined the recognizability and the fading of the original work in the new product (European Court of Justice, sentence from 07/29/2019 – C-476/17).

2.2 AI Act

The AI Act[2] aims to create a common legal framework for developing, marketing, and using artificial intelligence systems in the European Union (Recital 1 AI Act). The AI Act is based on a phased approach, characterized by the motto: The riskier an AI system, the more extensive the requirements it must meet.

As a product regulation law, the AI Act aims at providers and operators but also imposes obligations on those who import AI systems.

2.2.1 Risk-based approach

The AI Act differentiates between ’prohibited AI systems’, ’high-risk AI systems’, ’general-purpose AI models with systemic risk’, 'AI systems with moderate risk', and ’minimal-risk AI systems’. It focuses on the regulation of high-risk systems. These are AI systems that, due to their intended purpose, pose a high risk of harming the health and safety or fundamental rights of individuals by, among other things, significantly influencing the outcome of decision-making (Recital 52, Art. 6 paragraph 3 AI Act).

Article 16 of the AI Act gives an initial overview of the obligations and requirements for high-risk systems. The requirements include creating technical documentation, preparing error-free and complete training data, logging data during the system's life cycle, and ensuring an appropriate level of cybersecurity and robustness.

2.2.2 AI Act and GenAI

‘Large generative AI models are a typical example of a general-purpose AI model, as they allow flexible content generation, such as text, audio, image, or video content, which can easily contain a wide range of different tasks (GPAI)’. (AI Act ErwGr 99)

With this definition, the European Union attempts to describe GenAI systems such as ChatGPT or DALL-E. The AI Act further differentiates these ‘GPAI models’ by distinguishing between models with and without ‘systemic risks’ (Article 51, paragraph 1 of the AI Act). The European Commission decides whether there is a ‘systemic risk’. In addition, a presumption rule applies under which a GPAI model holds a ‘systemic risk’ when ‘the cumulative amount of computation used for its training measured in floating point operations is greater than 10(^25)’ (Article 51 paragraph 2 AI Act). Furthermore, providers of GPAI models are obliged to provide technical documentation for the model, including training and test procedures.

[2] https://eur-lex.europa.eu/legal-content/DE/TXT/HTML/?uri=OJ:L_202401689

It is particularly relevant that providers of GPAI models must prepare a detailed summary of capabilities and limitations to other AI providers integrating the GPAI model and publish information on training content used (Art. 53 paragraph. 1 lit. b), d) AI Act). Thus, the responsibility for the requirements of the AI Act along the 'AI value chain' is to be guaranteed (Art. 25 AI Act). This documentation also serves the purpose of transparency vis-à-vis authorities, which can request these documents on GPAI models (Art. 53 Abs. 1 lit. a AI Act).

The documentation and transparency obligations do not apply to GPAI models made available as a free and open-source license if there are no systemic risks (Art. 53 paragraph 2 AI Act). AI systems and models developed exclusively for research and development purposes are entirely exempt from the regulation (Art. 2 (6), (8) AI Act), provided they are not placed on the market or only put into operation for research purposes.

2.2.3 Schedule of the AI Act

The AI Act provides a phased timetable for the applicability of the regulations. It entered into force on 1 August 2024 with the publication in the Official Journal of the European Union. While prohibited AI systems must be taken out of operation after six months at the latest, the obligations for providers of GPAI models take effect 12 months after entry into force of the AI Act. Requirements and obligations for high-risk systems apply after 24 months at the earliest.

Companies that develop, offer, and operate AI should continuously review individual AI systems to determine their risk category. Caution is required when using GPAI models in high-risk areas (see Annex III AI Act). The competent authority (AI Bureau), set up by the Commission specifically for supervision and enforcement, will issue information and guidelines on dealing with the AI Act.

At the national level, responsible authorities must be appointed to monitor compliance with the AI Act. In Germany, this role is likely to be assumed by the Federal Network Agency. However, the necessary implementation law for the AI Act is still at the drafting stage.

2.3 Data protection and compliance

Currently, many products and services are available for different applications - often with the option of generating a limited amount of GenAI results free of charge. These services allow GenAI to be used without setting up its own infrastructure. They are, therefore, very convenient and, in many cases, can be used at a low threshold. In contrast to language models set up locally, language models hosted externally add essential aspects of data protection and compliance. As there is no data processing agreement and the exact use of the data is unclear, no personal data of employees, business partners, or other persons, and no confidential/secret information should be entered. This applies particularly to services free of charge.

2.4 License terms and conditions of use

Uncomplicated services with a low threshold tempt users to be too casual about the license conditions. As with any other type of software, the license terms and conditions of use of the model must be observed. A distinction is often made between private, research, and commercial use. Before using the software or API, it must be ensured that the use for the intended purpose is legally permitted.

2.5 Watermarks in GenAI results

It is negligent to hope that using GenAI will go unnoticed. Particularly in language models, the user can integrate unnoticeable watermarks. The GenAI can subtly manipulate the generated text so that the adjustments do not affect the text quality or content but remain algorithmically recognizable. Language models generate their content using so-called tokens, which follow a process of probability distribution. Watermark technologies can change this randomness in a certain way so that the next token is selected in a pseudo-randomized manner. It is not always clear whether service providers use watermarks and make use of possible recognition.

3 Major restrictions

Even though the results of GenAI are impressive and technical development is progressing rapidly, limitations still need to be considered.

3.1 Limited comprehension of context

Despite their considerable ability to 'understand' and generate language, language models cannot understand the real world, leading to potential inaccuracies or nonsensical responses. For example, the generated texts may be aimed at a different context and, therefore, be inappropriate for the specific context.

3.2 Lack of transparency

Due to their complexity and size, large language models act like a black box, making it difficult or even impossible to understand the reasons for specific results or decisions. Even if GenAI can partially explain itself, these explanations originate from a black box whose procedure usually remains inscrutable for humans. The basic mode of operation may be transparent. Still, the enormous size of the models means that individual answers cannot be understood or can only be understood with considerable effort.

3.3 Biases

Large language models trained with vast amounts of data can (unintentionally) adopt the biases contained in the source data. As a result, the models can generate potentially biased or distorted output. Therefore, when using GenAI where people are affected (e.g., personnel decisions), particular attention must be paid to such biases.

3.4 Hallucinations

GKI models generate new content based on what has been seen or learned. This is often clearly recognizable with artificially generated images, but not with generated texts. It is, therefore, essential to be aware that language models can freely invent answers. Even if a text sounds plausible and convincing, it may be a pure 'hallucination' of the GenAI. For this reason, great caution is required when using GenAI in critical applications. Critical decisions and information should not be based solely on a language model. This applies to potentially significant effects on people, lives, or health. Current GenAI models are very suitable for 'creative' applications but not yet for cases in which the factual correctness of the results is essential, such as in legal texts.

4 Clarification of functional requirements

Especially in professional contexts, GenAI must be used with due care and attention. Safe use requires checking the functional requirements of the system. In the following, we give an overview of essential points that must be considered.

4.1 Scalability and latency

The performance of a language model must be evaluated according to the requirements of the use case. Large language models require enormous resources (especially GPU memory), meaning only small batch sizes are usually possible when used locally. This means that the responses of the language model require more time. Therefore, the model's response time and scalability requirements should be considered early to ensure it can achieve the expected performance.

4.2 Interpretability

The results of language models are often neither interpretable nor comprehensible, even for experts, as the neural networks used are too large and complex. In specific applications, however, it can be essential to understand the model's decision-making process. This is particularly important in critical areas such as healthcare or the legal system. The use of language models must, therefore, be scrutinized critically.

4.3 Evaluation and Monitoring

The performance of a language model must be continuously monitored and evaluated to ensure that it continues to fulfill the functional requirements. In practice, this can be done using appropriate benchmark data sets, for example, to measure and evaluate its performance over time. This is particularly important when using APIs, as the underlying models can change and affect large providers.

4.4 Data protection and compliance

Language models can already gain access to sensitive information during the training process or process it during subsequent operations. Appropriate security precautions are therefore essential to restrict access to or transfer of sensitive data and protect privacy. This applies especially to the use of APIs and service providers.

5 Technical implementation and efficiency

GenAI models - even supposedly ‘small’ ones – require enormous resources. Depending on the planned application and functional requirements, local or externally operated instances can be more suitable. A thorough assessment must always be carried out regarding the prerequisites and requirements.

5.1 Using a local GPU instance

GPUs with large memory are recommended for local instances to load the largest possible GenAI models. While smaller models run on the system RAM and with a dedicated GPU or exclusively with a CPU, their performance (in terms of generated quality and response time) still needs to improve. Ideally, graphics cards with 80 GB VRAM are used for language models. Depending on the size of the language model and use case, the language model should be distributed across several GPUs so that larger batch sizes can be used. These systems usually cost tens of thousands to hundreds of thousands of euros.

5.2 Using an externally operated GPU instance

Language models can also be operated ‘classically’ on remote servers. High-performance hardware resources can be booked as required via external, globally operating service providers to ensure the hosting and scaling of large AI models. While the initial investment is very high for local instances, the costs per operating hour are comparatively higher for using externally operated cases. These are usually in the high single-digit euro range per GPU hour, so price and performance should be carefully compared in advance.

5.3 Using APIs

Besides the option of running the GenAI models on local or third-party hardware, there are also APIs available, usually fee-based. Three current, well-known APIs are (in random order):

- OpenAI API (https://platform.openai.com/docs/api-reference/introduction)

- Huggingface API (https://huggingface.co/inference-ap)

- Aleph Alpha Luminous (https://docs.aleph-alpha.com/docs/introduction/luminous/)

There are also other offers for manual use, such as the OpenAI Playground (https://chat.openai.com/) and other websites that use language models and are potentially suitable for minor manual queries, such as https://www.perplexity.ai/.

5.4 Ecological and economic considerations

Numerous calculation steps are required to generate content from GenAI models. This generally makes the use of GenAI extremely resource-intensive.

5.4.1 Energy requirement

GenAI models require enormous computing power and are highly resource-intensive. The training of large models takes place in large data centers with considerable energy consumption. Even after training, the energy requirement stays high, as each query involves many computing operations.

5.4.2 Hardware requirements

The creation and use of GenAI models lead to an increased demand for hardware, especially for powerful GPUs and servers. This is reflected in the price of this hardware.

5.4.3 Data requirements

The creation and training of GenAI models require enormous amounts of data to achieve suitable results. This data must be collected, stored, and made usable. This step also requires technical and human resources.

6 Summary and checklist

This white paper summarizes the key challenges and aspects that must be considered when using GenAI. GenAI can and will revolutionize work. However, its use requires careful consideration and planning.

If you think about using GenAI, the following checklist containing the most essential points to be clarified in advance might help you:

- Conceptual suitability for the application
  - ☑ Does GenAI make sense for the application?
  - ☑ Are there less complex models and algorithms available that would be more suitable?
  - ☑ Are explainability and comprehensibility essential criteria?
  - ☑ May the available data be legally used for GenAI?
- Criticality of the applications
  - ☑ What is the level of criticality of the application?
  - ☑ What further decisions are affected?
  - ☑ Should liability be considered for the decisions?
  - ☑ How should possible reservations be handled?
  - ☑ What safeguards will be required?
- Technical resources
  - ☑ Are there local resources to operate the planned GenAI models?
  - ☑ If not, is a corresponding investment possible?
  - ☑ Can the use of external hardware or services be a viable alternative?
- Licenses and legal issues
  - ☑ Are the models intended for the planned application/project/operation?
  - ☑ What license requirements or terms of use must be observed?

This checklist is not exhaustive and does not fully cover all relevant aspects. However, it should serve as a starting point for planning and implementing a profitable use of GenAI.

If you are considering using GenAI in your company, please do not hesitate to contact us.

Disclaimer: Some of the content of this white paper was created using GenAI. This applies to the introduction and the title graphic.

7 Change log

- Version 1.0: Initial publication
- Version 1.1: Supplement of copyright aspects and requirements from the AI Act
- Version 1.2: Update to the AI Act

8 Contact

Dr.-Ing. Fabian Rigoll

+49 721 9654-552

rigoll@fzi.de

FZI HQ Karlsruhe

Dr.-Ing. Steffen Thoma

+49 721 9654-840

thoma@fzi.de

FZI HQ Karlsruhe

Maria Rill, M. Sc.

+49 721 9654-646

m.rill@fzi.de

FZI HQ Karlsruhe

FZI Research Center for Information Technology

Haid-und-Neu-Str. 10–14

76131 Karlsruhe

Germany

+49 721 9654-0

fzi@fzi.de

www.fzi.de/en

EDIH AICS.
EUROPEAN DIGITAL INNOVATION HUB
Artificial Intelligence & CyberSecurity

Co-funded by the European Union

W4.0 Initiative Wirtschaft 4.0 BW

digitalLÄND

Baden-Württemberg Ministerium des Inneren, für Digitalisierung und Kommunen

Baden-Württemberg Ministerium für Wirtschaft, Arbeit und Tourismus