Sherlock Holmes in Orbit

Artificial Intelligence detects
Biotopes on satellite images aided by
synthetic Data Sets

Download of this Slide Deck:
EN: sherlock-ai.climatehackerz.com
DE: sherlock-ki.climatehackerz.com

Jörg Osarek, May 2023
m:contact at climatehackerz.com
T: +49-151-23-0-24-333

2023 - 2035

About Joerg Osarek

- Born 1970 in Bad Homburg, Germany
- **Industrial Business Management Assistant** (IHK) 1991 Hewlett-Packard
- **IT-Consulting** since 1992
- Founder: Business consultancy 2003,
- **Co-founder its-people** group, 4 years CEO at its-people Hochtaunus GmbH
- Founder: Gordon's Arcade publishing business 2012
- http://skilltower.com/publications.html (60+ records)
- YAM-Track-Record since the 1980s*
  *YAM = Years ahead of Mainstream
- **Founder of ClimateHackerz.com 2019**

About Joerg Osarek: Published Books

2023: Work in Progress

Travelguide.ClimateHackerz.com

2008

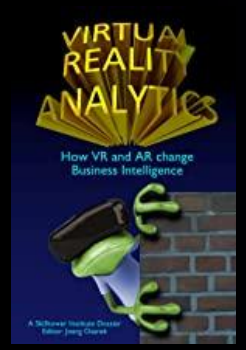
2016

2012

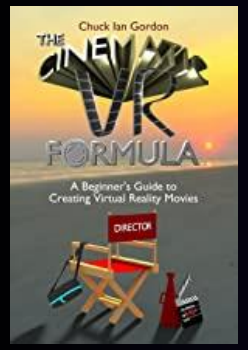
2014

2012

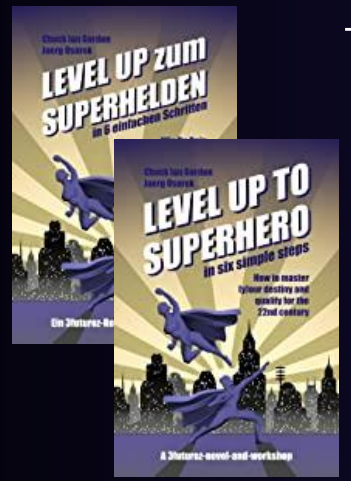
2016



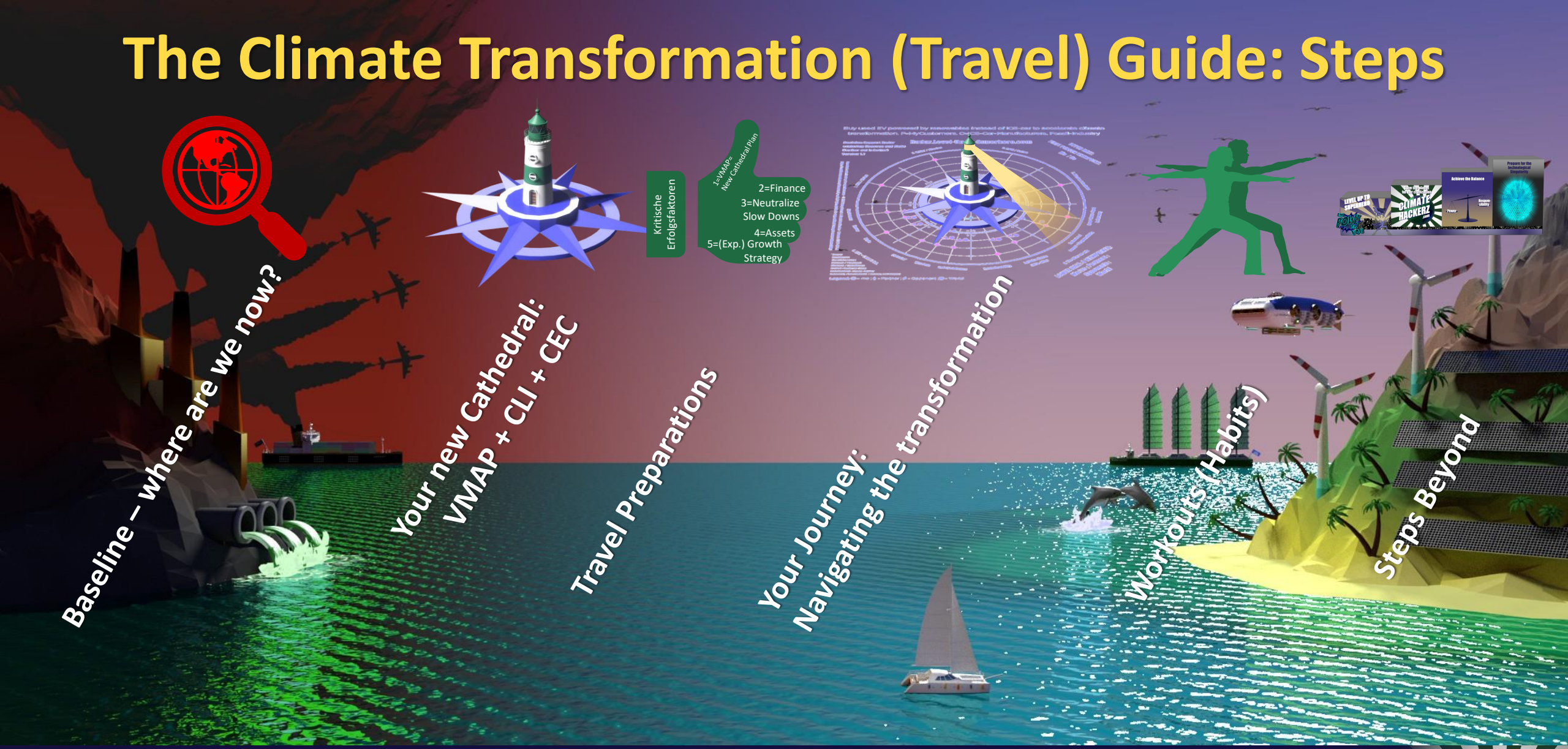
2017



2020

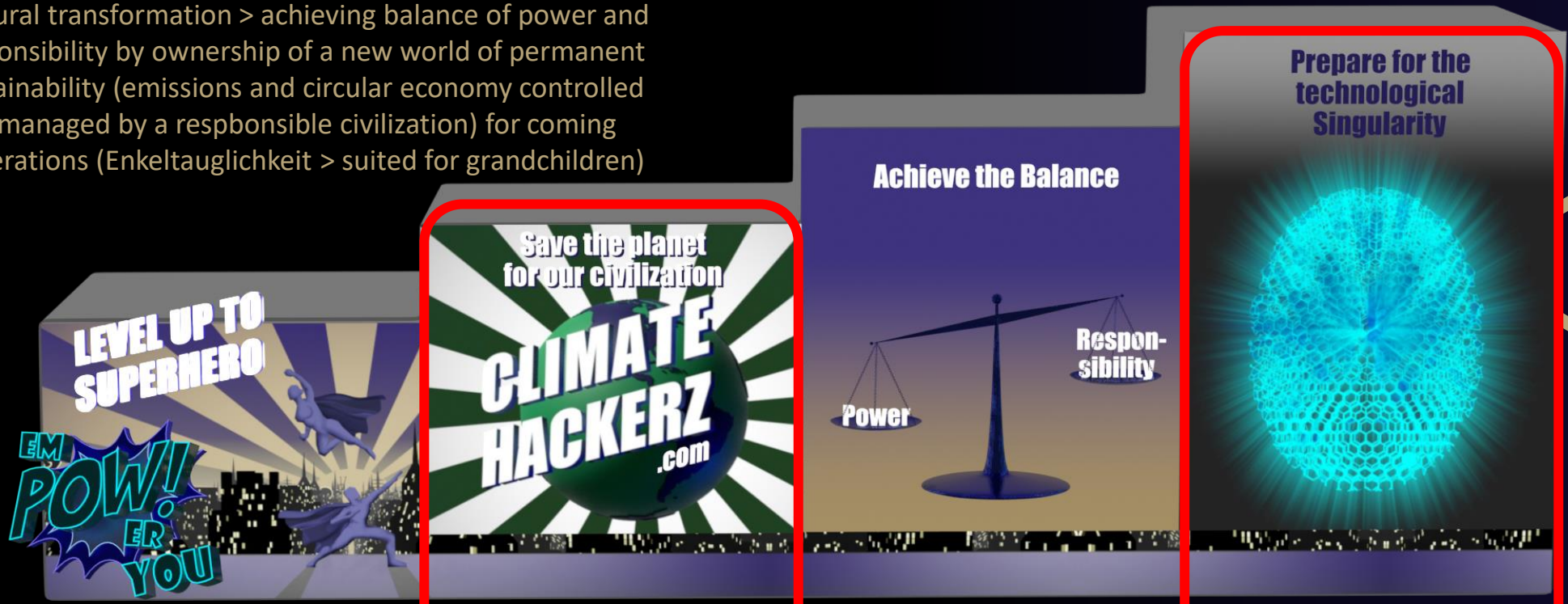


The Climate Transformation (Travel) Guide: Steps



The climate transformation is only one of the grow-up-tests our civilization currently faces

Cultural transformation > achieving balance of power and responsibility by ownership of a new world of permanent sustainability (emissions and circular economy controlled and managed by a respbonsible civilization) for coming generations (Enkeltauglichkeit > suited for grandchildren)



Level Up 1: Empower our character and orgs

Level Up 2: Solve climate crisis with Doughnut economics (by Kate Raworth)++

Level Up 3: Bring balance to our civilization

Level Up 4: Singularity preparation management **(AI/ML revolution)**

succeed → Q22C = Chance for Utopia

fail → If we don‘t do enough = guaranteed dystopia = hell on earth for most people.

www.Q22Century.org

Sherlock Holmes im Orbit

Artificial Intelligence detects Biotopes on satellite images
aided by synthetic Data Sets

on a celebratory occasion like this,
I thought I'd perform this a little poetically.

WHY? – the Kickoff

From CSRD to Green Deal we aim
For sustainability without deforestation's shame
trained eyes must map biotopes: intact or lost
But mappers are scarce and come at high cost.

Thus, AI's demand has surged to new height
For mass evaluation, it's become a sight
Millions of times we'll turn to AI's trained skill
To find habitats on sat images, as easy as to chill.

What? – The task

A simple task, it seemed to be
Name biotopes with AI's clarity
Like Sherlock Holmes in orbit above
To designate with precision, no need to shove.

To know how this system can work
Where science and tech stand, a perk
and scaling up was what we aspired.
A proof of concept was required.

What we learned from this POC
I share with pleasure with you ad hoc.

POC – What did we learn?

We took Hadden's investment rule as an advice.

Why only build one if you can have two at twice the price?

Two vendors, we chose, from business and from research

interrogating them for the answers we search

With challenge and cross-fertilization, we put them to test

Got clarity for our path, and we got to know them best.

POC – What did we learn?

AI experiments were the hands-on approach we took,

A blessing for our team, to learn and be part of the hook.

Our project became a journey of discovery and play,

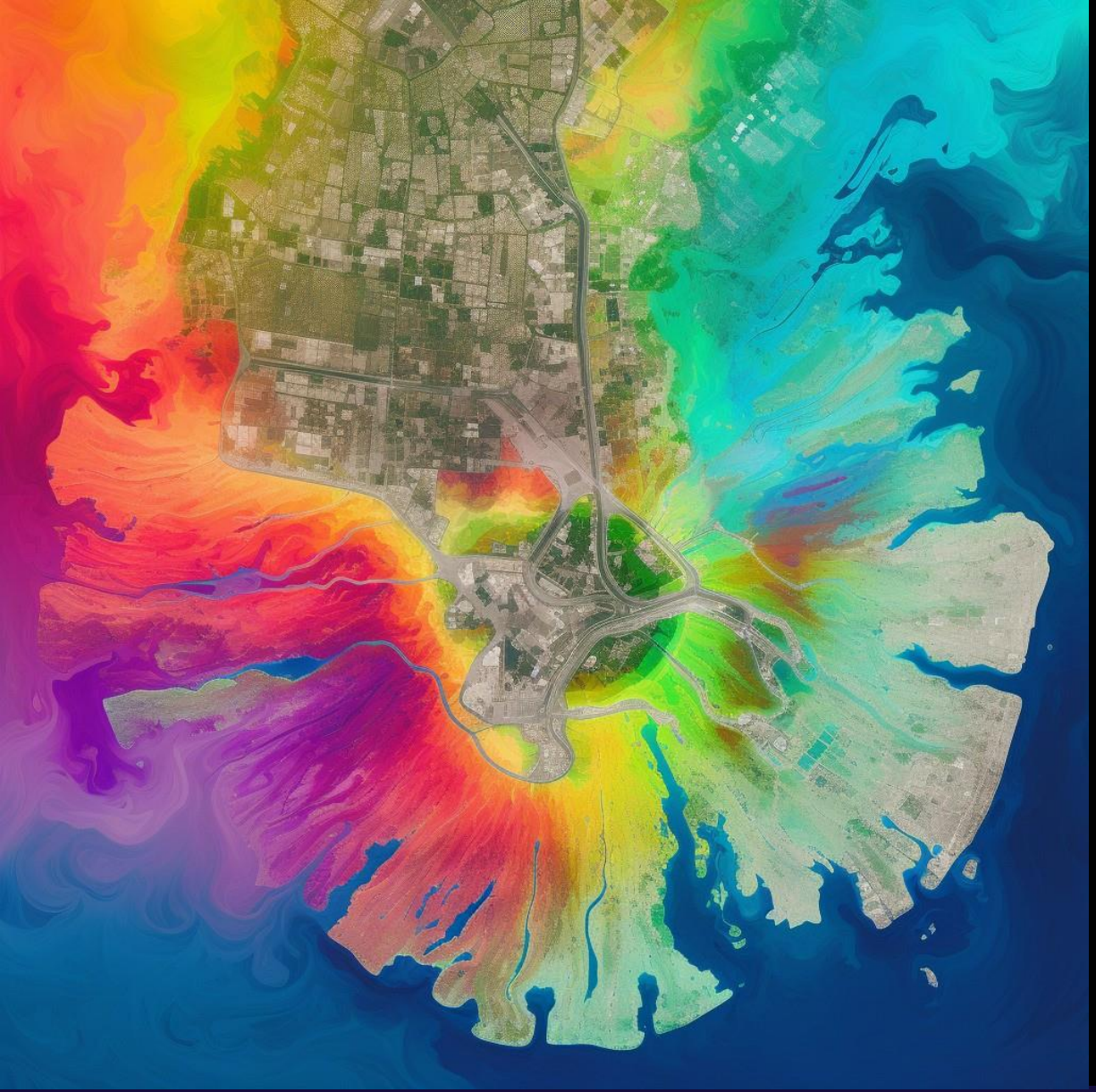
Better results together, in a harmonious array.

POC – What did we learn?

Picture in, answer out, a wish so fine,

But in reality, it's not that simple to align.

Core idea: create an open and flexible AI toolbox for a variety of tasks.

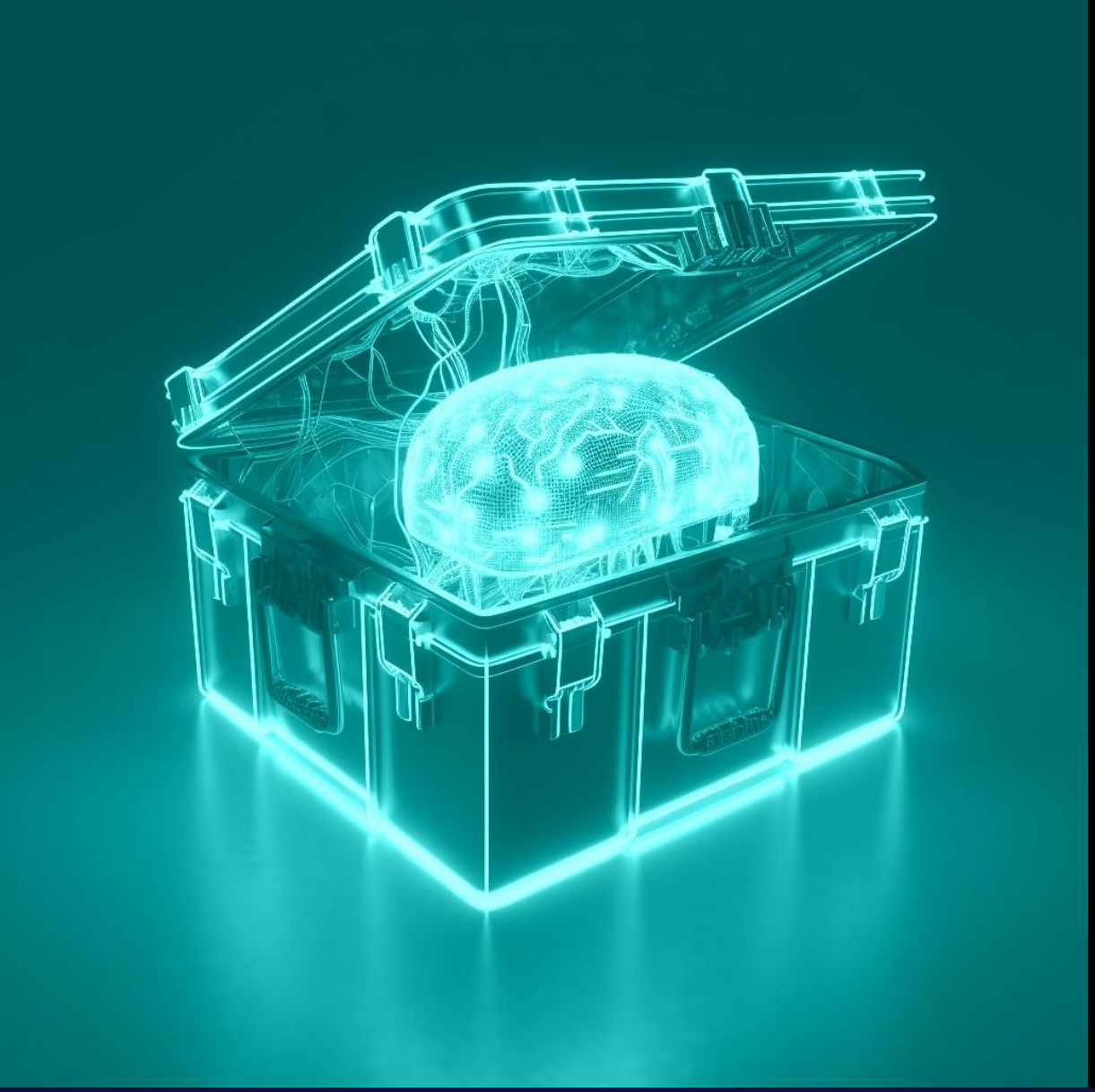


POC – What did we learn?

Explainable AI offers answers in a blink.

XAI lets you watch the AI think.

Core idea: create an open and flexible AI toolbox for a variety of tasks.
* important especially now in times of approaching danger by crises and the proximating singularity: explainable AI - XAI. Quote from - https://www.ibm.com/watson/explainable-ai : (Explainable artificial intelligence (XAI) is a set of processes and methods that allows human users to comprehend and trust the results and output created by machine learning algorithms. Explainable AI is used to describe an AI model, its expected impact and potential)



POC – What did we learn?

How to build such a system - an important debate.

a modular IT architecture we need to create.

A flexible toolbox for many configurations

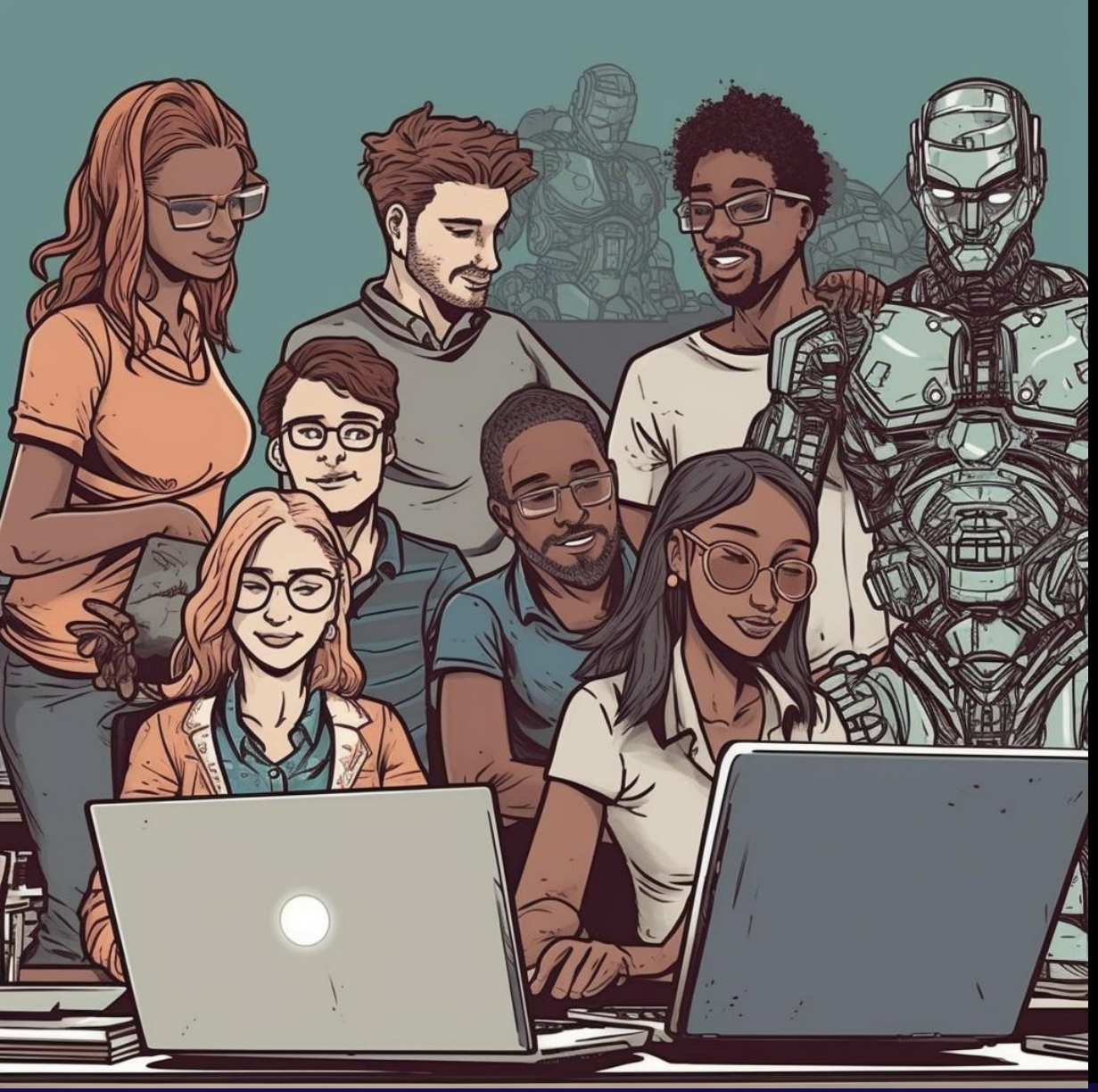
adaptable to a lot of situations.

In turbulent accelerating times with a lot affected

only by flexibility investments are protected.

Core idea: create an open and flexible AI toolbox for a variety of tasks.

POC – What did we learn?



Finally, an enthusiastic interdisciplinary team is most important to achieve that dream.

Core idea: create an open and flexible AI toolbox for a variety of tasks.

POC – What did we learn?

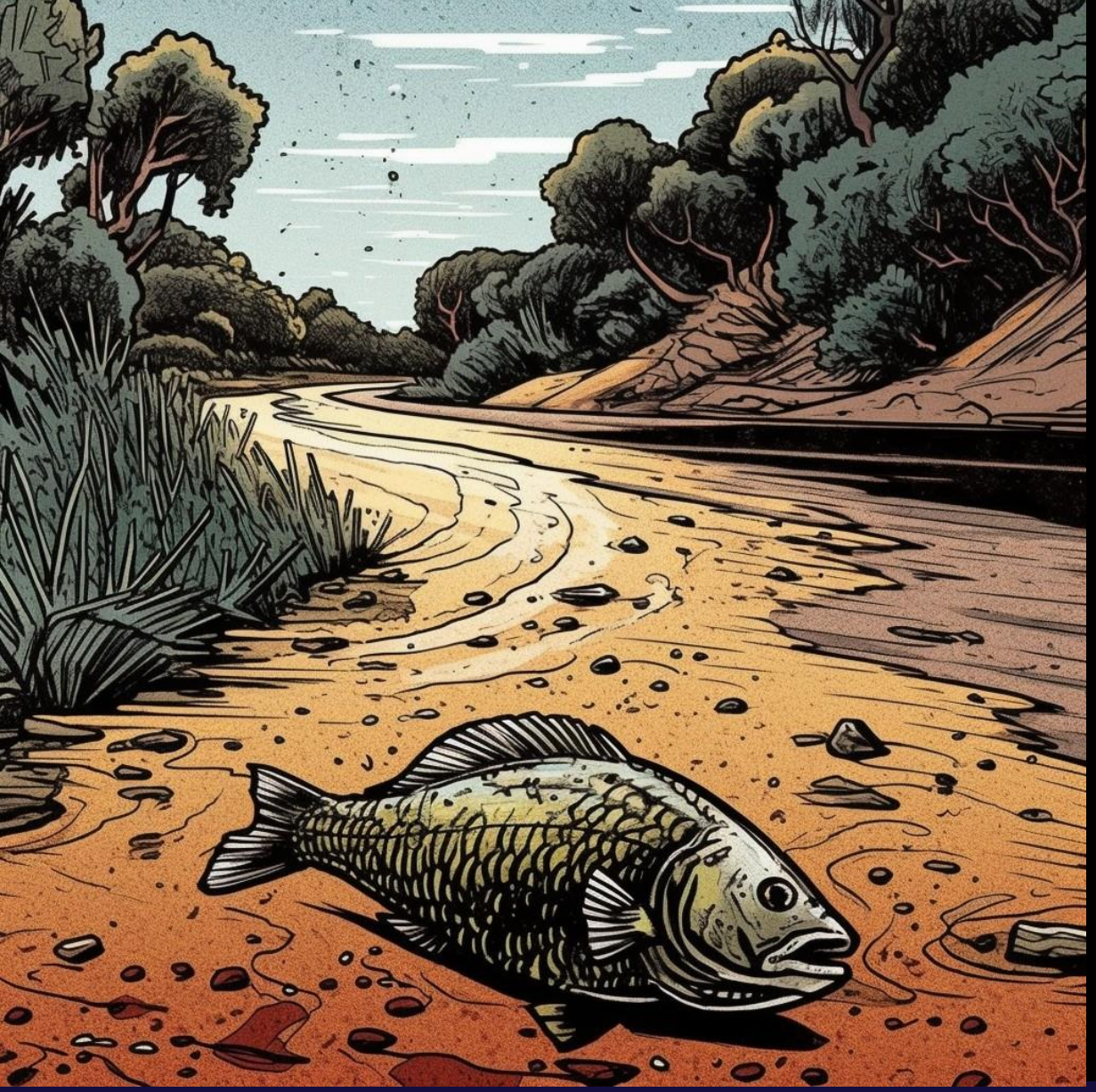
We fell in traps, what a mishap!

How did we master them and adapt?

POC – What did we learn?

The watershed segmentation confused tree with shadow.

Forest recognized but it was only the meadow.



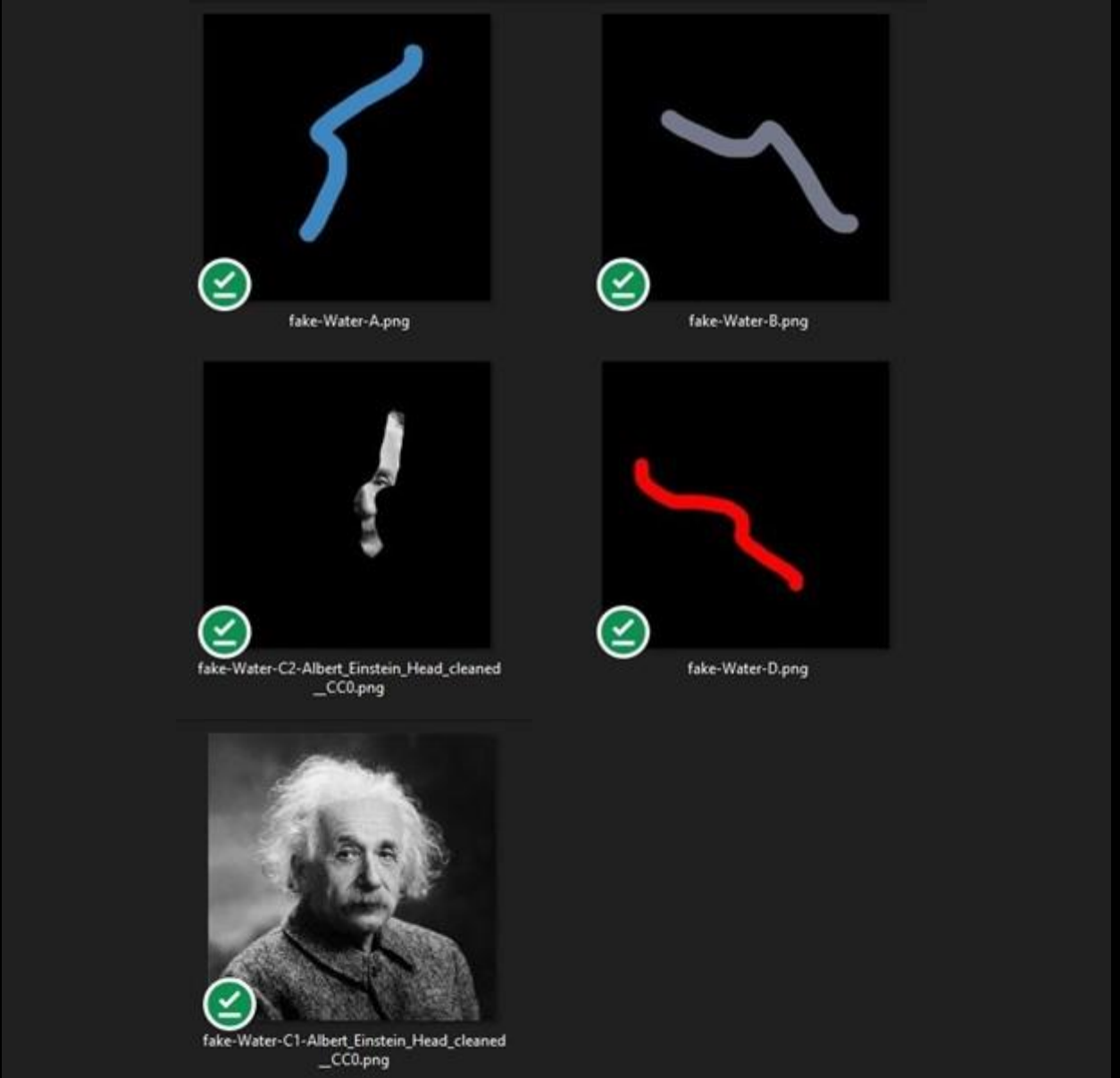
POC – What did we learn?

To achieve high accuracy, you need to be aware precise geo projection is where you have to take care.

By 7 meters the system sometimes missed.
You think that's not much? Here comes the twist.

A fish in a creek can perfectly thrive.
But next to it on the hiking trail it won't survive.

This can be solved with strong test management to fix flaws that - thanks to XAI - we can comprehend.



POC - What did we learn?

Curiosity uncovered a bias in prediction,
Too many watercourses, an erroneous depiction.

Elongated shapes, not only the blue,
were identified as water, no matter the hue.

Einstein's photo, an unfortunate case,
Classified as watercourse, a misplaced trace.

Einstein's whole photo classified as scree - an illusion
By only training with Biotopes, we ourselves caused this confusion.

The AI confused signal with noise.
To include "Others" in the training may be our next choice.

Lessons Learned: Pitfalls and solution approaches
Problem: Bias recognizes long shapes as water. This can be avoided by changing the training strategy, because it can be e.g. also paths or buildings.Images of rivers with insufficient intellectual creation by Jörg Osarek
Image of Einstein - Photograph by Oren Jack Turner, Princeton, N.J. Public domain - https://commons.wikimedia.org/wiki/File:Albert_Einstein_1947.jpg

POC - What did we learn?

biotope recognition was partly good, partly a farce.

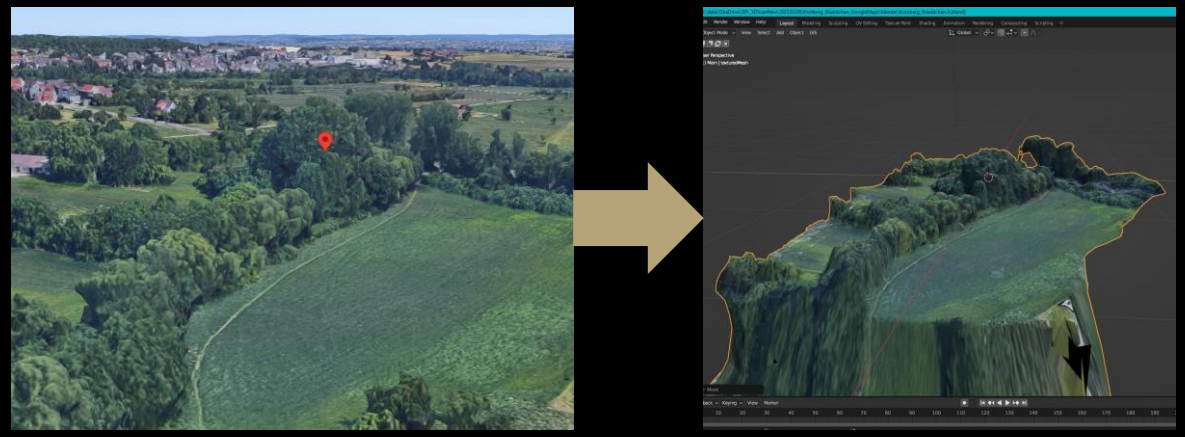
The amount of training data was simply too sparse.

Affordable human labeling could not be completed.

A blow to the business case - are we shamefully defeated?

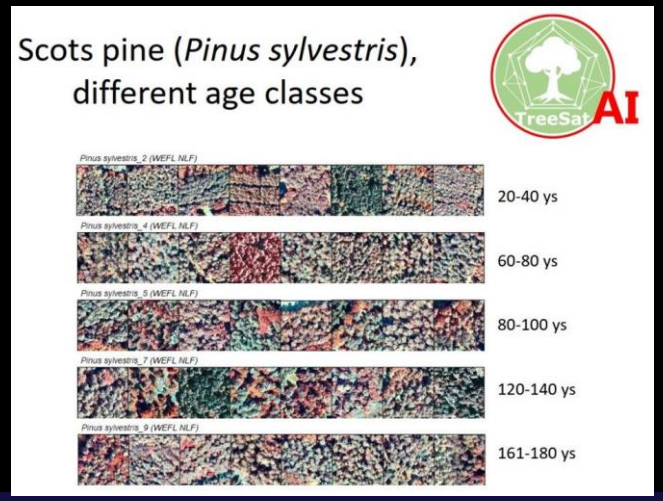
Problem: The amount of training data was insufficient. But labelling by humans led the business case „ad absurdum“ as we wanted to achieve affordable classification in the first place.

POC - What did we learn?



3D elevation models aid one part of the task,

More spectral bands are another vital ask.



TreeSat AI Study:
https://www.tu.berlin/geoinformation/forschung/projekte/laufende-projekte/treesatai
und
https://zenodo.org/record/6598391#.ZCLBFXZByUk – letzte Seite des PDFs:

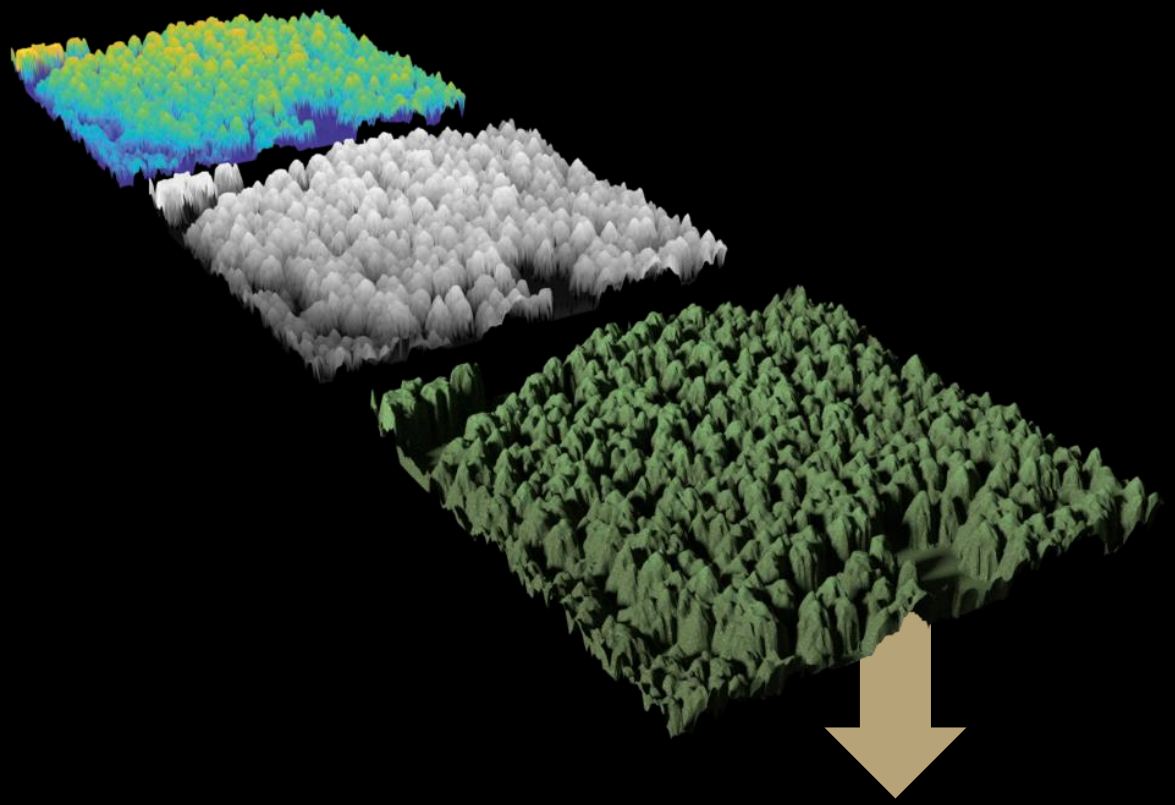
The TreeSatAI study proved what was sealed:

Using Infrared, the age of pine trees can be revealed.

Problem: The amount of training data was insufficient.

* 3D Elevation models – top left Google Maps Screenshot > top right 3D Model in Blender created with photogrammetry – small forest close to Kronberg Ts. Germany
- Sat Radar Canopy height model = obtain more information channels and train the AI with them additionally.
- Additional spectral bands possible like NIR, Infrared, see Study TreeSat AI (link above)

POC - What did we learn?



The ML model identified this as forest.

How do we obtain that training data just debated?

Some from Satellites, some can be self fabricated.

Synthetic data sets is the method's name.

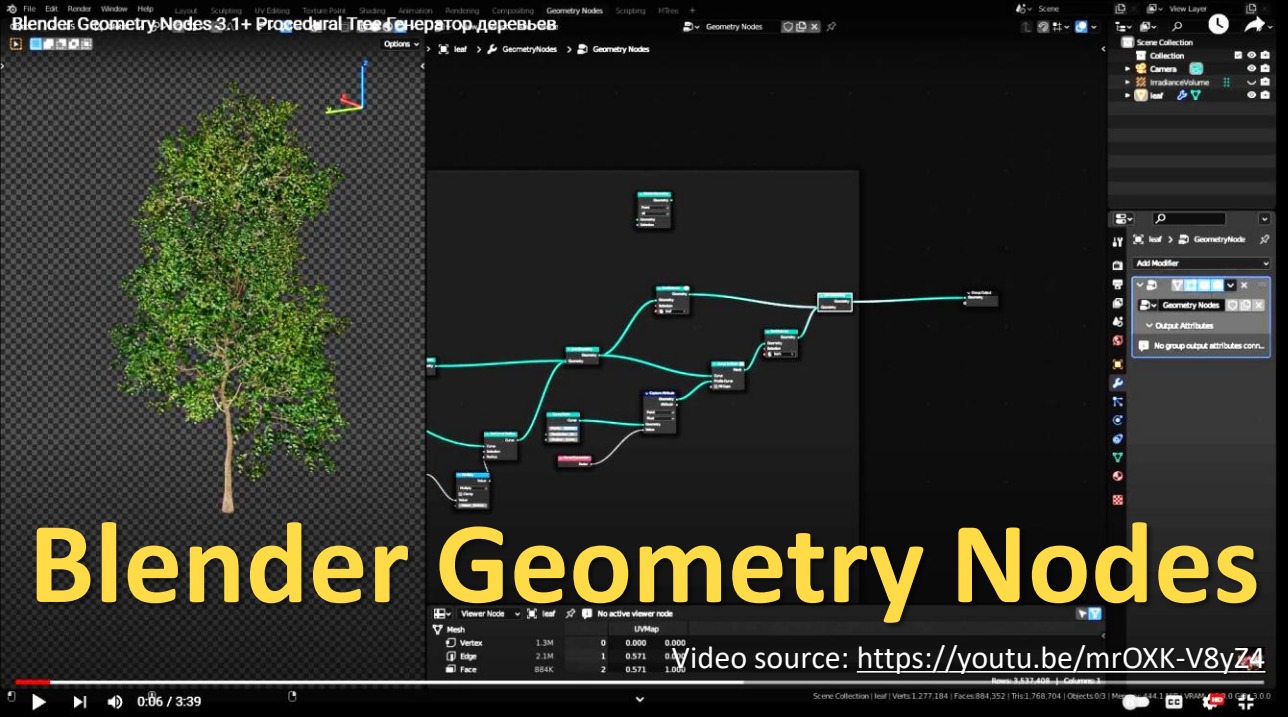
For autonomous driving it already earned some fame.

But is it suitable for biotopes as well?

We proved AI recognition for them as I can tell.

Problem: The amount of training data was insufficient.
Idea: synthetic Data Sets to the rescue > a proven 3d data set was built that was accepted by the AI and identified as forest.
Illustration © by Joerg Osarek – 3d model and lighting created in blender

POC - What did we learn?



Video source: https://youtu.be/mrOXK-V8yZ4

The free 3D Software blender can generate as we found

procedural vegetation in an infinite amount.

Automatically labeled, the data masses can grow

training will cost much less, we want you to know.

Multispectral and visible light, both we need,

To train AI well, and let it succeed.

Deep 3D know how is needed I have to say

But this doesn't scare the 3D enthusiast away.

Blender Addon Botaniq –

Images: https://blendermarket.com/products/botaniq-trees

Problem: The amount of training data was insufficient.

* Blender Geometry Nodes can be used to procedurally generate and simulate complex 3D scenarios. For vegetation, there are now Geometry Nodes based plugins, some of which are available for a fee, to build an arbitrary Synthetic Data Set pipeline that can provide optical and multispectral images as well as other data such as labels, and height information - as needed for the specific task. We have demonstrated experimentally: The AI accepts these synthetically generated images and recognizes forests from the near-photorealistic rendered image.

POC - What did we learn?

Midjourney Autumn forest
Seed: 1359038443

Midjourney Spring forest
Seed: 17121117

If with synthetic training data you fiddle around
maybe help from Generative AI can also be found.

Images are created with text prompts that we tweak.
Lately new achievements are announced each week.

Can such GANs help for training on biotopes as well?
We are just investigating so that soon we can tell.

It was amazing how with that POC our understanding grew.
I hope there is also something useful for you.

If you have any questions, feel free to ask,
I'm always happy to help, that's my task.

Problem: The amount of training data was insufficient.
* It is unclear today whether GANs (Generative adversarial network) on the market can be controlled with sufficient precision over a longer period of time to generate sufficient fidelity to reality. If one trains such a GAN oneself and provides it only with accurate, real and relevant training data, there should be no reason why such a GAN should not be able to precisely generate endless biotope images. It seems promising to split the training pipeline into 1. a first generative part (Synthetic Data Sets), 2. training a special GAN with it and 3. generating the synthetic datasets promptly based (Text to Image) directly with this trained GAN and 4. training the actual neural network with the synthetic datasets generated by the GAN. It will be helpful not only to generate images, but also depth maps and other spectra at the same time. Note: This theory has to be proven. If this proves to be viable, it can be used to create very effective and efficient AI training data pipelines for highly complex biological, geological, or even man-made environments, as well as for wildlife (fauna) or even medicine. Big advantage: I have the construction of the training pipeline completely under control (like the climate in a greenhouse) and the training pipeline for the respective ML project is thus significantly simplified for the ML trainer.

Next Step: A scaled system

An automated system monitoring the global biosphere
Wouldn't that be brilliant and awesome here?

To bring such a noble system to scale
Investment is needed for us to set sail.

Our POC final report has 47 pages.
Busy investors would read for ages.
How could I simplify
to explain it on the fly?
I'll try a poem on famous stages.

Illustration from „Die Sterntaler“ by Viktor Paul Mohn, 1882 (1842-1911), Public Domain – copyright expired – https://commons.wikimedia.org/wiki/File:%D0%92-%D0%9F-%D0%9C%D0%BE%D0%BD-%D0%97%D0%B2%D0%B5%D0%B7%D0%B4%D0%BD%D1%8B%D0%B5_%D1%82%D0%B0%D0%BB%D0%B5%D1%80%D1%8B.jpg

Sherlock Holmes in Orbit:

Artificial Intelligence detects Biotopes on satellite images aided by synthetic Data Sets

Download of this Slide Deck:
EN: sherlock-ai.climatehackerz.com
DE: sherlock-ki.climatehackerz.com

Jörg Osarek, May 2023
m:contact at climatehackerz.com
T: +49-151-23-0-24-333

2023 - 2035