Decker File Edit Card Tool View Widgets Help

What does it mean to 'generate' information?

[work in progress] PAS, November 2023

Some motivations

- Information-talk everywhere in CogSci. But not enough focus on the *operations* on informational states.
- Neural information-processing lingo: selection, integration, amplification, attenuation… ***generation*** [some exceptions: Fazekas & Nanay 2021; Fazekas & Overgaard 2018]

- Old philosophical chestnut - how's **new** knowledge possible? (Plato + footnotes and footsoldiers; Brit-empiricists: how is **abstracting or generalizing** possible)

- And knowledge is made of information… right?

→ The goal: **provide a definition of what is information-generation for a cognitive (esp. neural) system to contribute to this chestnut.**

Some background

- Inference as **‘filling-in’** the gaps. Think inferential statistics (vs. descriptive)
- In perception: **ecological vs inferential** [active, constructive]; filter-out vs construct from the **gist**
- ‘Passive’ theories: **all’s ‘in the signal’.** To perceive: just make sense of the bombardment!
- Internal representation ***detached* from input** (visualization, counterfactuals, internal models, recall etc.). They have to contribute as a source of information!

- Generative AI; Rumour has it: hot and new capacity. But what exactly is this generative component, really? [**following Buckner’s 2018 *modus operandi***]
- Some other debates: e.g. consciousness studies (Kanai et al. 2019); concepts (Quilty-Dunn 2020)
- Recent .RAR Wars in AI research community about the explanatory purchase of compression.

Maybe it's already settled?

- Loss of information is generally understood in algorithmic information theory terms, i.e. defined over real patterns.

- But what about the informational ***gain***? Lossy operations are not reversible!
- Gain is usually equalled with mutual information or KL-divergence. (Cover & Thomas 2006)
- But that's not generation. Neither it satisfies us in philosophical explanation what does it mean to gain new information.

Pump it up and weed it out

- Intuition pump: **to generate is to (at least) create new information.**
- But clearly **just churning new strings that merely eat up (memory, channel) space is different from truly *new information*** being produced. Think of 10k string of random signs vs 10k 0s.
- **Neither does simply adding new symbols or degrees of freedom.** Neurogenesis does not equal to an informational brain-gain. Right?
- Transforming **just what's 'already in there' does not qualify as generative, either**. Otherwise, it's trivial. Every new output would be 'generated'.
- Different cases of ***faux*-generation** we need to winnow out: copying, translation (but: geometric vs. linguistic), integration, downsizing. Proc-gen?
- But: let's **not overmentalize either**, we want naturalistic explanations with AI systems as thinking tools, proof-of-concepts or models.

Garden variety after weeding out (mini-recap)

- How are these faux-generative processes different from truly generative ones we hunt down?

- How does new information emerge that **is not already entailed** in the available or accessible in the data?
- We need a **‘contentful’** understanding of generativity.

Generative models

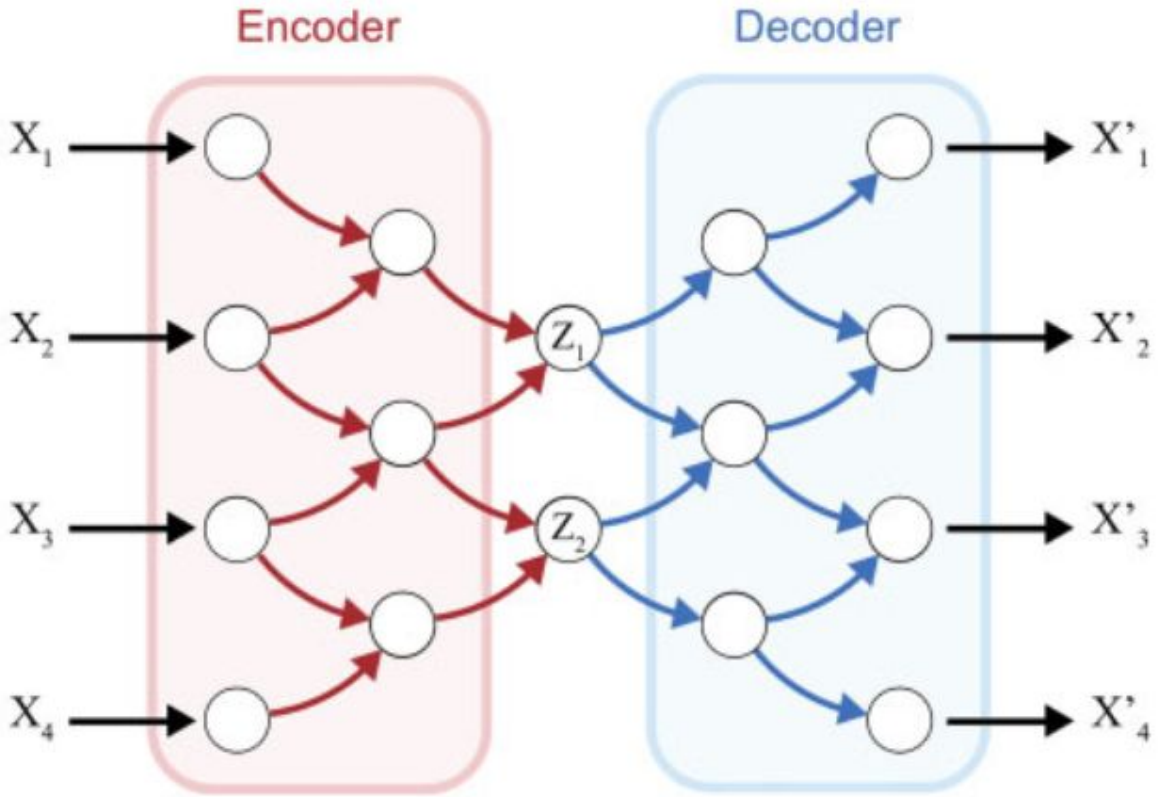
- Generative model—**statistical model of the process of how sensory data are generated** from a set of hidden causes.

- **Functionally different** from more classical discriminative models.
- **Explanatorily crucial** in predictive processing - top-down predictive architectures. Radicals: all content is predictive and comes from generative processing (Clark 2015). ‘Hypotheses’ or ‘hyperparameters’ as generators?
- But also semantic pointers! (Thagard & Stewart 2014; Eliasmith 2013; Blouw et al. 2015)

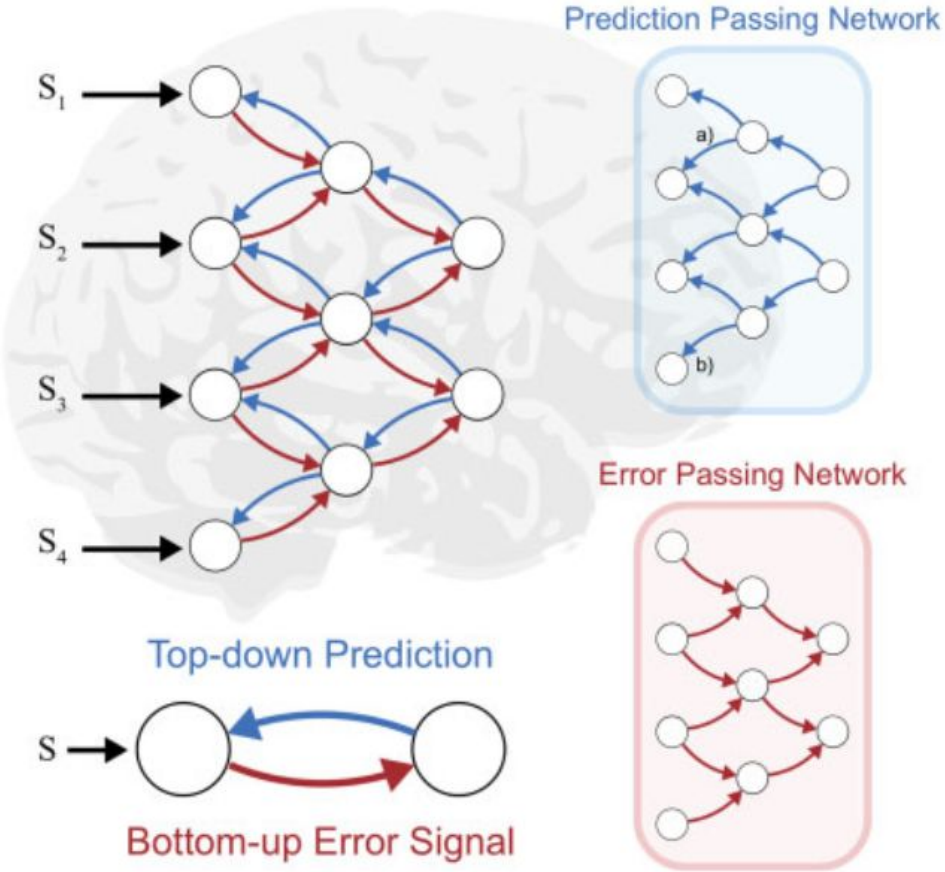
Autoencoders as an example

Autoencoders work by *encoding* unlabeled data into a **compressed representation,** and then *decoding* the data back into its original form. “Plain” autoencoders were used for a variety of purposes, including reconstructing corrupted or blurry images. *Variational* autoencoders added the critical ability to not just reconstruct data, but to output variations on the original data. (Martineau 2023)

(a) Autoencoder



(**b**) Predictive Coding



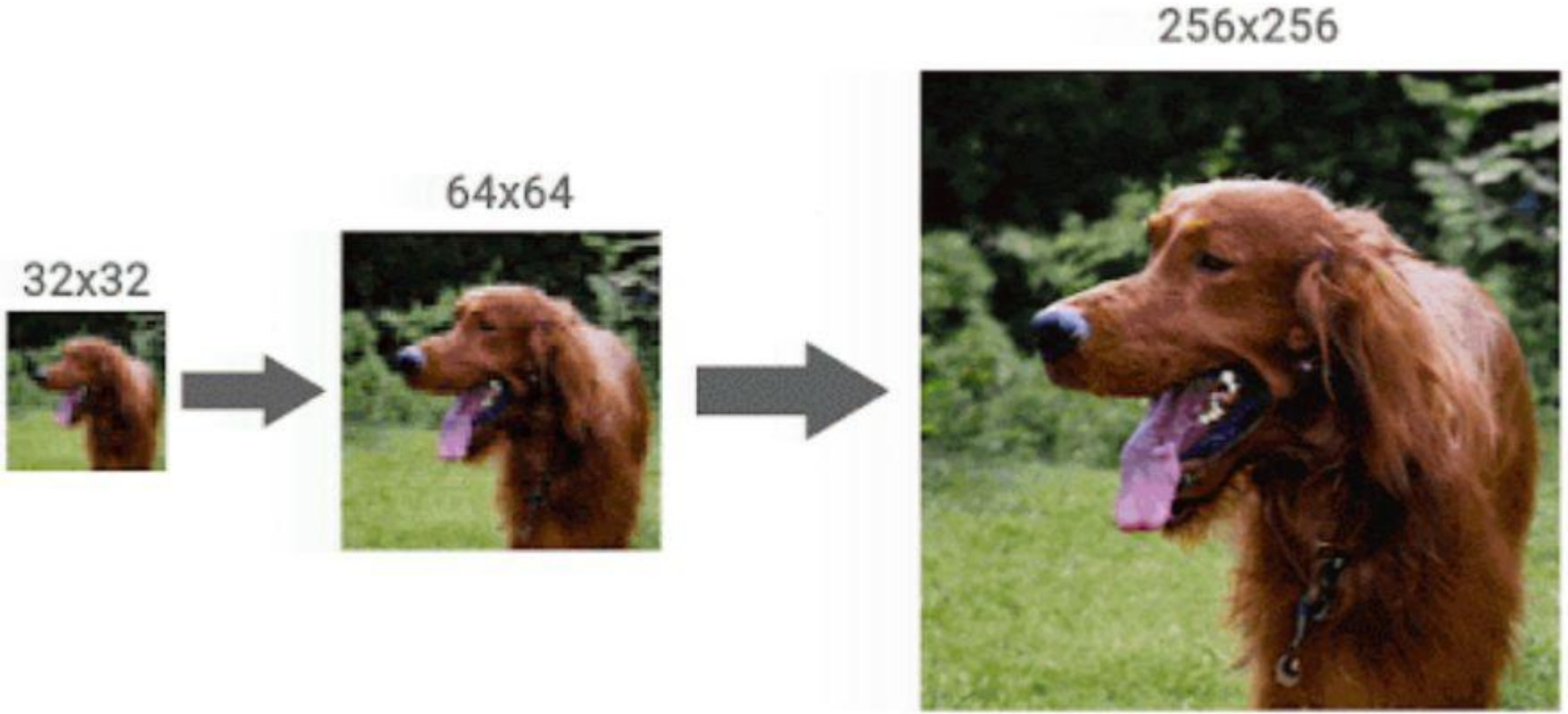
Ok but so what?

- We argue that information generation can be seen as production of sensory representations using internal models. This is achieved by ***reversing*** **the process of representation learning**, which is projection from sensory inputs to internal models. (Kanai et al. 2019:4)

- So it is a **decompression from the latent space**.

- But it's **not a regular decompression** as it as merely an expansion of what's already there.
- Think **zipfiles** then compare to **upscaling**. Operating solely on the file itself only gets you so far. You need to mobilize the background knowledge.

Upscaling example



Toy examples - computational photography

- Let's take a step back & use computational photography **as a sandbox to create thought experiments** in order to adjudicate cases via philosophical intuitions.
- Computational photography defined as 'computational imaging techniques that enhance or extend the capabilities of digital photography in which the output is an ordinary photograph, but **one that could not have been taken by a traditional camera**' (Mark Levoy; after Zubaryev 2020)
- **Generated = what is not actively stored 1:1 all the time**
- Modern photography is filled to the brim with these clever tricks and enhancements (stacking, plenoptics) - **does it render all photography generative now?**

In the side-by-side above, I hope you can appreciate that Samsung is leveraging an AI model to put craters and other details on places which were just a blurry mess. **And I have to stress this**: there's a difference between additional processing a la super-resolution, when multiple frames are combined to recover detail which would otherwise be lost, and this, where you have a specific AI model trained on a set of moon images, in order to recognize the moon and slap on the moon texture on it (when there is no detail to recover in the first place, as in this experiment). This is not the same kind of processing that is done when you're zooming into something else, when those multiple exposures and different data from each frame account to something. This is specific to the moon.

CONCLUSION

The moon pictures from Samsung are fake. Samsung's marketing is deceptive. It is adding detail where there is none (in this experiment, it was intentionally removed). In this article, they mention multi-frames, multi-exposures, but the reality is, it's AI doing most of the work, not the optics, the optics aren't capable of resolving the detail that you see. Since the moon is tidally locked to the Earth, it's very easy to train your model on other moon images and just slap that texture when a moon-like thing is detected.

Now, Samsung does say "No image overlaying or texture effects are applied when taking a photo, because that would cause similar objects to share the same texture patterns if an object detection were to be confused by the Scene Optimizer.", which might be technically true - you're not applying any texture if you have an AI model that applies the texture as a part of the process, but in reality and without all the tech jargon, that's that's happening. It's a texture of the moon.

According to a report by Zhihu (via Android Authority), there could be some shenanigans going here. It appears that Huawei is using AI to do more than just enhance pictures of the moon taken by the P30 Pro. Zhihu's Wang Yue has run some tests and has come to the conclusion that rather than merely enhancing a photo of the moon, Moon Mode takes images of the moon already available, and inserts them into the user's photograph. Yue took a number of photos to support his theory, using Moon Mode to snap pictures of other subjects besides the moon. His conclusion is that the Moon Mode does not produce totally original photographs.

Futurism
23 godz. ·

It goes beyond being a gaudy photo filter to something much more insidious and lifelike.

futurism.com

Google's New Phone Adds Smiles to Your Photos Even If Everyone Was Miserable

The Dua Lipa Conundrum

Let's consider a **thought experiment** - we give the computational pipeline a task and it outputs us the very same pic on the right (100% pixel values equivalence)

- **Case A** - imagine a search engine query 'albanian pop-singer'
  - searches the database and brings back the relevant target from the memory space
- **Case B** - imagine a prompt to genAI 'generate an Albanian pop-singer'
  - creates it from scratch
  - uses a more general representation and the result is partly spurious, but not exactly accidental (process is optimized for more generalized task); would still count as boosting performance
- **Case C** - imagine a prompt to genAI 'generate a pic of Dua Lipa'
  - it uses the trained representation of actual DL photos to come up with an 'unseen' (not yet existing) one

- **Case** * - imagine a prompt to genAI: 'generate the best picture possible'

Still The Albanian Case

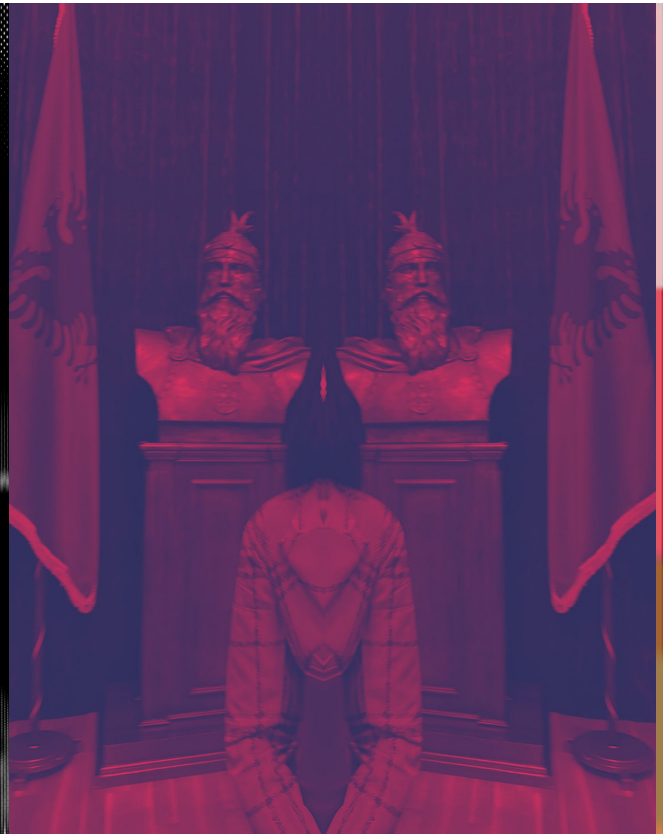
- **Case D** - **reverse deformation**: imagine we have a heavily transformed original picture and we want to go back to the original form but we do not store the history of transformations, we just want to use the ‘gist’ plus known general statistical dependencies
- **Case E** - **reconstruct (predict; guess?) from non-causal seed** - imagine a similar task, but where we try to get a DL picture from a random seed; e.g. prompted with: ‘how would this blob look if it was Dua Lipa’
    - note: in the example on next slide I did produce the maximally distorted version ‘causally’, but for the sake of the illustration let’s imagine it is a more or less random blob

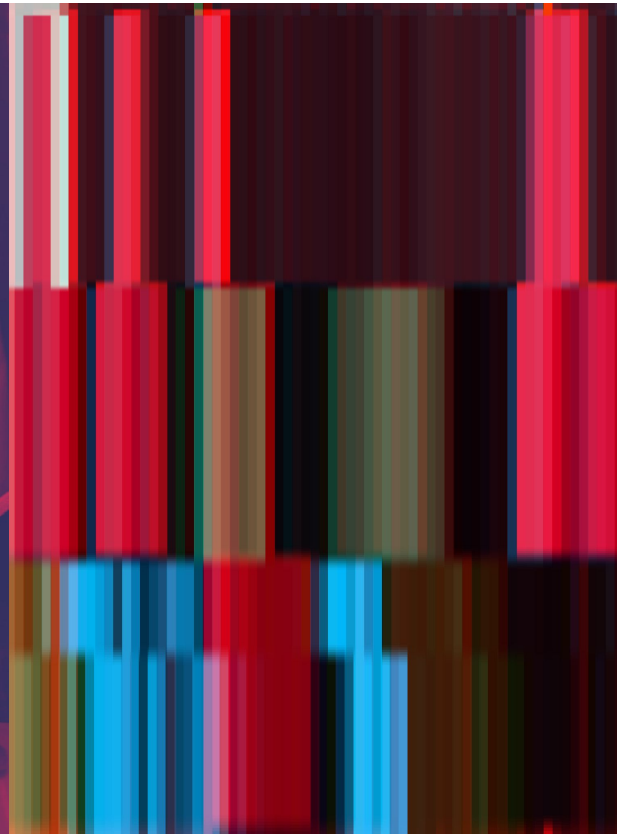
- In algorithmic information theory the limits of **reconstructability are in theory upper-bounded by the compressibility. Loss of information is non-reversible**. In theory to arrive at the same (original) picture you need information input (assumptions + generation or storage **coming from outside to fill in**)
- How are all these cases different? Are they on a spectrum? Or it’s just a family of distinct processes? **When it is merely transformed and when generativity kicks in?**

Limited distortion (Case D lite)

More lossy distortion features lost? - (case D hard)

Almost a seed (Case E)





Ways out

- Ill-posed problem
- These are separate computational processes not along a single dimension that just yield equal results.
- **There is come continuum and we need a boundary!**

Let’s not fret and probe this option.

Interpolation vs extrapolation

- Intuitively: **interpolation is ‘filling in’ the missing values that fall ‘between’ already known** datapoints, whereas

- Extrapolation is a (**statistically) educated guess about some value outside the ‘convex hull’** of the state space delineated by known data.

- Usually used to operationalize the ability of the models to **generalize well**.

- It seems to capture well the idea put forward here between **true ‘generation’** and mere ‘**unfurling’ or ‘expansion’** of already known information.

Inter/extra: does the dichotomy help?

- However, data scientists are recently **skeptical that for high-dim spaces (dim>100), the dichotomy is a real one: as true interpolation seems extremely unlikely**. (Balestriero et al. 2021)

- Trouble? Perhaps it is a case of data analysis not mapping well to true psychological dimensions and as such - bad candidate for explanations (results in very abstract spaces can be artifactual or epiphenomenal) for real cognitive processing (analogously to the Burnston-Shea (2021) debate on PCAs impact on the ontology of neural vehicles) [also: Bowers et al. 2022; Ritchie et al. 2018]

Some objections

- Is it a pseudoproblem then?
- However, aren't all these just mathematical functions (albeit fancy) that map input to output [**transforms**]?
  - Stochastic parrots (E. Bender)
  - One big jpg (T. Chiang)
- How to differentiate biased processes from legitimate inferential generativity? (Orlandi 2014)
  - Real informational differences, not 'explanatory gloss' (Cao 2020)
- **Are these good *explanations* of *cognitive* phenomena?**

The challenge

- From the standpoint of algorithmic theory of information what's 'extendable' (intuitively: unfurled) from the regularities is not really new information.
- Any such 'filling-in' according to patterns **is less informative** than random noise generation (or even not at all?).
- But it clearly seems like generativity both intuition-wise and on account provided here. How so?

- Two conceptual ways out: 1) **brute generation =/= actual generativity**; 2) generativity can be assessed for the system only if the relevant information encoding a given regularity that the system acts upon **is actually represented both *in* and *for* the system)**. In other word: in the virtue of its content.
- Let's combine them!

The gist

- Proposal: previously stored information mobilized for generation of a new batch of information (variation on Boghossian's 'taking condition'?)
    - How it's implemented - tbd. (weights. connections. other types of memory - e.g. molecular; Colaco 2022)
- This stored information has to contentfully correspond to some regularities that enable that **filling-in** (as in: gist+completion paradigms: fuzzy-trace theory type (Reyna & Brainerd 1995; in AI research: decompression from a seed sampled from more latent, low-dim space)

- **No generation without compressed information,**
    - **New information tokens are generated if and only if: previously stored information about underlying regularities that are explicitly represented for and accessed by the system are used in the virtue of their content.**

Semi-conclusions - 3 desiderata towards a definition

- Lost (or not had in the first place) information has to come from somewhere - back to the idea of inference as filling in the gaps

Desiderata:

1. **Takes advantage of regularities** nonetheless (prediction is based on the underlying structure) - just **actively represented and used** -> **storage** inside **compressed representations**
2. Internally stored information is mobilized to perform the task -> **decompressed** (from low-dim latent space) **to generate new tokens.**
3. Production is performed ***in the virtue*** **of informational content.**

How to do this mapping? That's the actual filling-in job [to do].

- The goal: dissect the network's parts (functional units) to **show only certain architectures achieve generativity through the realization of these theoretical desiderata.**
- Where and how these regularities are **actively stored**
- How are they acted upon to **systematically transform the signal in a regimented way according to these represented regularities.**
- Boundary problem vs manifold problem - solutions has to have answers for these
- How the end-product is truly new (generative), not in the sense of just being outside the training regime, but as **contentfully** generative.

Thank you for your attention