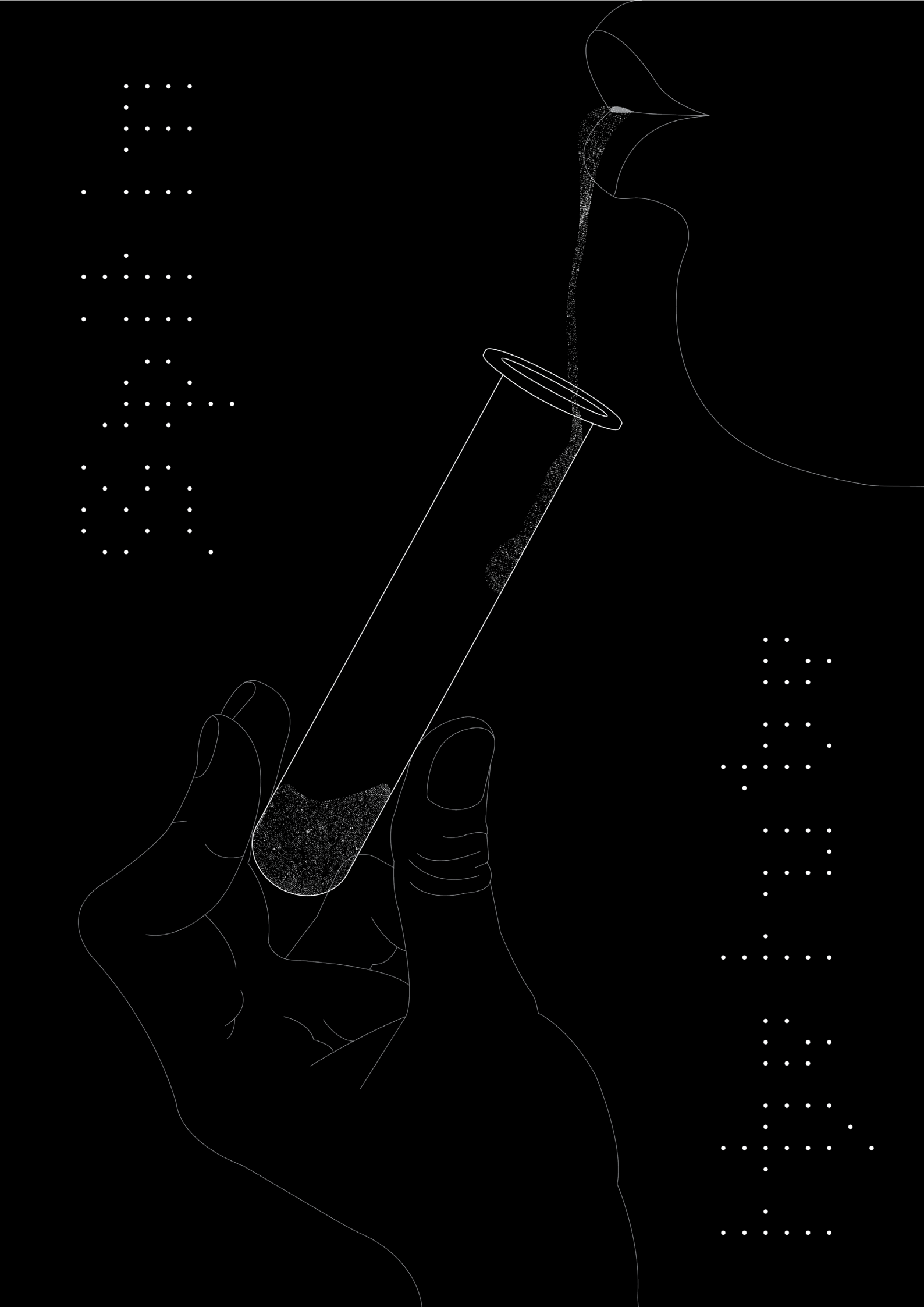


SPIT IN THE TUBE

DECEMBER
2021

Authors:
Andrej Petrovski
Olga Tanasković, PhD / Biology
Jelena Adamović, LLM / Legal

Editor:
Milica Jovanović

Illustrations and design:
Kristina Pavlak

SHARE
FOUNDATION

CONTENT

I was spending some time in the US in September 2017, after a conference I was supposed to be speaking at in Miami got canceled due to Hurricane Irma. With a lot of free time to meet friends and roam around NYC, I got into this mood of extensive online shopping, mostly books and gadgets, and eventually an Ancestry DNA kit. The exact reason for this purchase remains unknown to me to this day, since I had no doubts regarding my "lineage" (though, one never knows what their ancestors might have done in the past). Let's say it was an impulsive purchase for pure entertainment. And that's how I sold a copy of my DNA to a corporation, for pure entertainment.


GENESIS

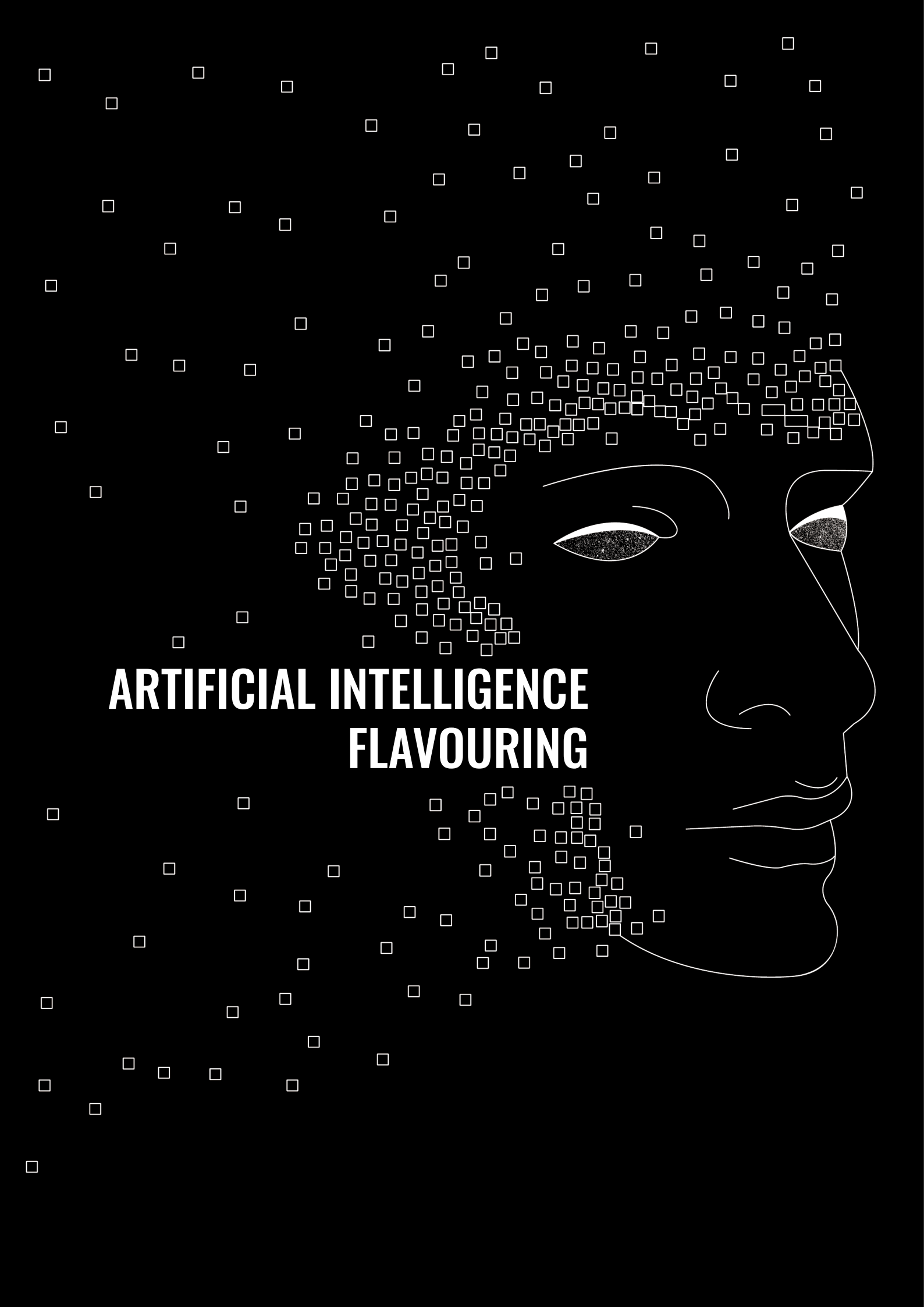
The Genesis, or creation of life, is a fascinating phenomenon. People have, for ages, tried to attribute it to a Divine being, for a simple reason of not being able to understand how life is created. On a meta level, even nowadays most of us cannot apprehend the concept entirely. Whether it might be for reasons of mere curiosity, or to clarify events that have happened in the past, many of us have at some point thought that digging through our genealogy might provide answers which were otherwise inaccessible.

In the past, families had kept a family tree adding each new generation over time. Many take pride in their origins and keep those family trees to prove that somehow their origin makes them better than others.

On the other side, the evolution of technology has quantified society to the smallest particles. From tracing users' behaviour and shopping preferences, monitoring telecommunication metadata, we have come to facial recognition and biometric surveillance. The relationship between the corporations which profit from processing personal data and the state apparatus which conducts mass surveillance has already been established as an unbreakable bond from which both sides profit in their own way; companies by earning a lot of money and states by controlling society.

Telecommunications, especially with the introduction of 5G and implementing network interfaces to lightbulbs, refrigerators, dishwashers and cars, or what they call the Internet of Things, has served as a great tool for both profit and control. However, what corporations and states couldn't tap into is making an interface out of the human body. And that's when biometric mass surveillance was born. Its predecessors, smart watches and fit bands, still needed a device to be attached to the human body, whereas biometric surveillance eliminates the intermediary on the side of the human and captures data whether or not there is a device attached to the body.

Face recognition has become the first form of biometric surveillance that has gone mainstream, as both fingerprinting and iris scanning require some level of cooperation from the surveilled individual. Across the world this technology has been implemented in various settings, from Walmart, to Social Credit Systems in China, from football stadiums in Denmark and other EU countries, to the entire city of Belgrade, Serbia. How this technology was developed can to great extent tell us how some day our own DNA might be used as a surveillance parameter.


ARTIFICIAL INTELLIGENCE
FLAVOURING

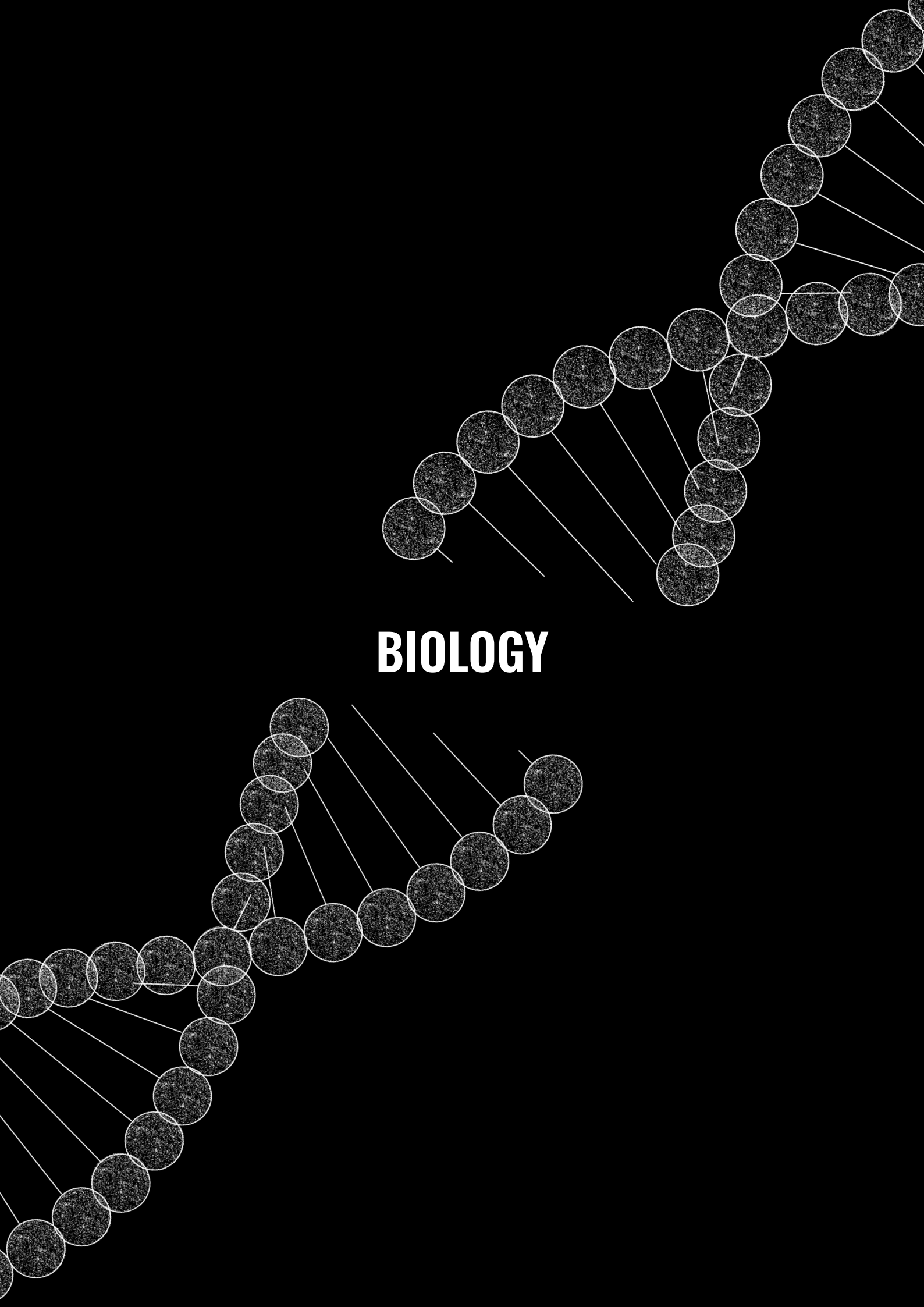
The principal requirement for biometric surveillance is technology which has the ability to quantify biometric characteristics of the human body. In the case of face recognition, the technology needs to be able to detect human facial features, quantify and store them in a database. For analytical purposes, quantified data are compared and the similarities between two images of faces are determined. If the similarities are above a certain threshold, then the algorithm decides that both images are of the same face. Now, this is an oversimplified description of a Face Recognition system. However, in order to train an artificial intelligence algorithm a training set of images of faces is necessary.

In most cases, for training purposes, the so-called MegaFace dataset is used. This dataset is one of the most important benchmarks for commercial vendors and it contains almost five million faces from over three million photos downloaded from almost 50 thousand Flickr users' photo albums. State authorities, on the other side, have datasets that consist of biometric photographs of citizens taken for the purpose of issuing travel documents or other forms of photo ID documents.

When training an AI algorithm, the key aspect is the balance between precision and high level of detection abilities. If the AI is too precise, it will be rigid and images of faces that either have bad quality or were taken from a bad angle, won't even be considered as faces. On the other hand, if the AI is too liberal, it will match faces that do not represent the same person, more often than not. In a recursive process of Machine Learning, the AI is tweaked until a balance is reached for an acceptable functionality. This process is extremely approximative and arbitrary, and currently is the most common way an AI algorithm is trained. Enlarging the training dataset does to some extent improve the functionality, but at the end of the day we will be still dealing with a limited dataset and human bias in terms of determining the threshold of acceptable functionality. This brings us to a conclusion that such systems can never, given the current possibilities, be free of bias and highly accurate, yet all-catching, i.e. this is one of the systems that cannot really be improved greatly.

How do we draw a parallel with what's happening now with faces and DNA? Well, at the moment there are multiple private companies around the world which are building datasets that can be used for Machine Learning purposes for algorithms which in the future will be able to perform biometric mass surveillance using DNA as an input.

Telecommunication surveillance is possible to bypass by leaving all devices at home, even techniques to sabotage surveillance capitalism exist, but when it comes to biometrics, it's much harder to avoid being a part of the surveilled area, since it's really not possible to leave your face at home. Makeup and masks can help, until the AI is not fed with a training set to surpass those obstacles as well. However, we cannot change our DNA, since that's the essence of who we are.

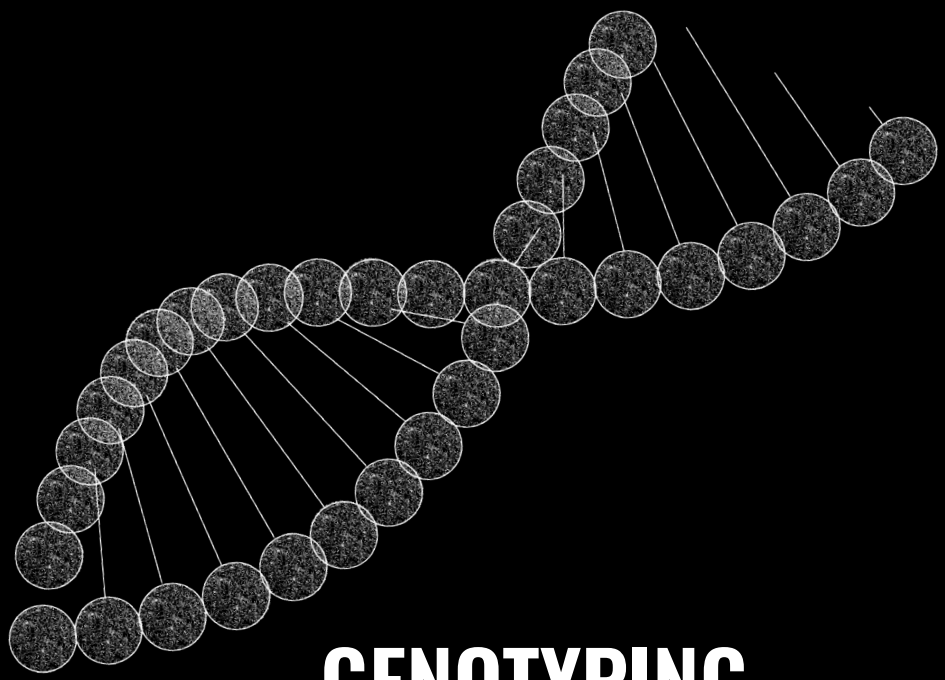


BIOLOGY

DNA is the code of life, a molecule in a form of a double-helix that stores genetic information about every living organism on Earth. This genetic information is represented by the sequence made of four chemical bases: adenine (A), guanine (G), cytosine (C) and thymine (T). These bases always pair together, and along with phosphate groups and sugar bind the two complementary DNA strands. This pairing always occurs in the same way - A pairs with T, C with G. The specific sequence of these four letters in our genome, which is about 2.9 billion base pairs long, conveys vast amounts of information. This is similar to the way a series of zeros and ones tell a computer how to run a complex program, or to alphabet letters that, when put in a certain order, form words and sentences.

We get half of our DNA from each of our biological parents, and we will pass on a selection of half of it to our children. Human DNA is 99.9% identical from person to person. However, there are small differences, called variants (or single-nucleotide polymorphisms, SNPs) that make each of us unique. Exactly these variants are being looked at when performing commercial DNA tests to find more about our ancestry or genetic conditions.

In the last two decades, the technology that allows us to read our DNA has improved dramatically and the cost has lowered down significantly, which allowed a lot of companies to produce their own commercial DNA kits. Today, there are companies such as 23andMe, Ancestry, MyHeritage, etc, that sell DNA kits to anyone who wants to find out about their ancestry or predisposition for developing certain health conditions. Apart from these implications, some other start-ups (GenePlaza) emerged on the market, which offer to reveal whether you have the possibility to taste cilantro as soap, or to know whether you have a “gay” gene.

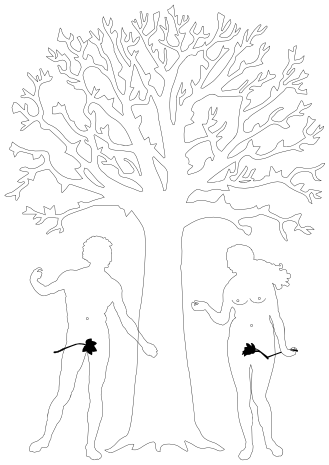


GENOTYPING

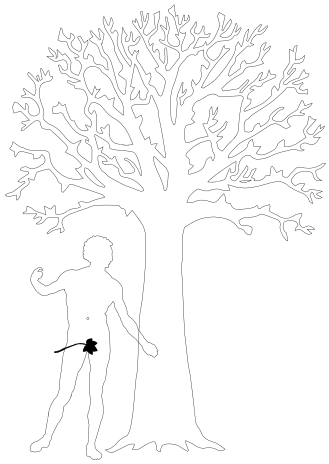
For genealogy analysis, all of these companies perform genotyping as a technique to detect specific variations in the customer's DNA. We can consider these variations (SNPs) as typos in our "book of life". SNPs occur in our genome every 200 - 1000 bases and are conserved during evolution, thus they can provide a unique fingerprint for use in identity testing.

Depending on the DNA, there are several subtypes of genotyping:

AUTOSOMAL DNA TEST



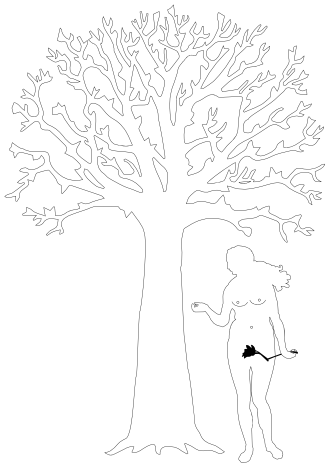
May be taken by either a male or female. Nearly everyone inherits 23 pairs of chromosomes from their parents. X and Y are sex chromosomes, while the rest of them (22 pairs) are autosomal (autosomes). Autosomal and sex chromosomes together contain a total ~20 000 genes. These genes are 99.9% identical in all human beings, but the small variations (SNP) are the ones that make the difference among all of us and are responsible for our inherited traits and conditions. This testing can provide you with the information about your ancestry and chances of inheriting certain conditions.



Y DNA TEST

Can help to determine the common parental ancestor. Since the Y chromosome is passed down from fathers to sons almost completely unchanged, this test can be taken only by a male. The Y-DNA line can be traced hundreds, even thousands years back. Completing a Y-DNA test is therefore beneficial when you have hundreds of autosomal DNA matches but are having difficulty determining who belongs on your paternal line of interest. Y-DNA test results will also provide many matches, but they will all belong to that same Y ancestral line.

MITOCHONDRIAL DNA TEST (MTDNA)



Follows the maternal line. mtDNA differs from autosomal DNA and Y-DNA, whose genetic data is located in the 23 pairs of chromosomes residing within the cells' nucleus. Mitochondrial DNA is found outside the nucleus, inside the organelles called mitochondria which are responsible for producing energy for each cell. mtDNA is inherited only from the mother line, therefore both men and women carry their mother's mtDNA, but only women will pass that mtDNA down to the next generation. Since mtDNA is highly conserved, an mtDNA test can reveal that two people are maternally related. However, due to the high similarity, it cannot reveal their relationship, so those with matching mtDNA may be mother and daughter, sisters, aunt and niece, or very distant cousins.

Genotyping looks for information at a specific site in our DNA, where researchers already know that the important data can be found. Although important for the research, this method allows us to read only < 0.02 % of our total genome. On the other hand, with next generation sequencing (NGS), the exact sequence of a certain length of DNA can be determined. It can be a short piece, the whole genome, or parts of the genome such as the "exome" which are the regions of the genome that contain the instructions for RNAs (ribonucleic acid) and proteins. Depending on the region, a given stretch of sequence may include some DNA that varies between individuals, in addition to regions that are constant. Thus, sequencing can be used to genotype someone for known variants, as well as identify variants that may be unique to that person. For people with specific needs (undiagnosed medical issues, for example), sequencing is becoming a more and more attractive option. At the moment, there are some companies that have employed this method, such as DanteLabs, Nebula, and Ancestry (for their health options tests).

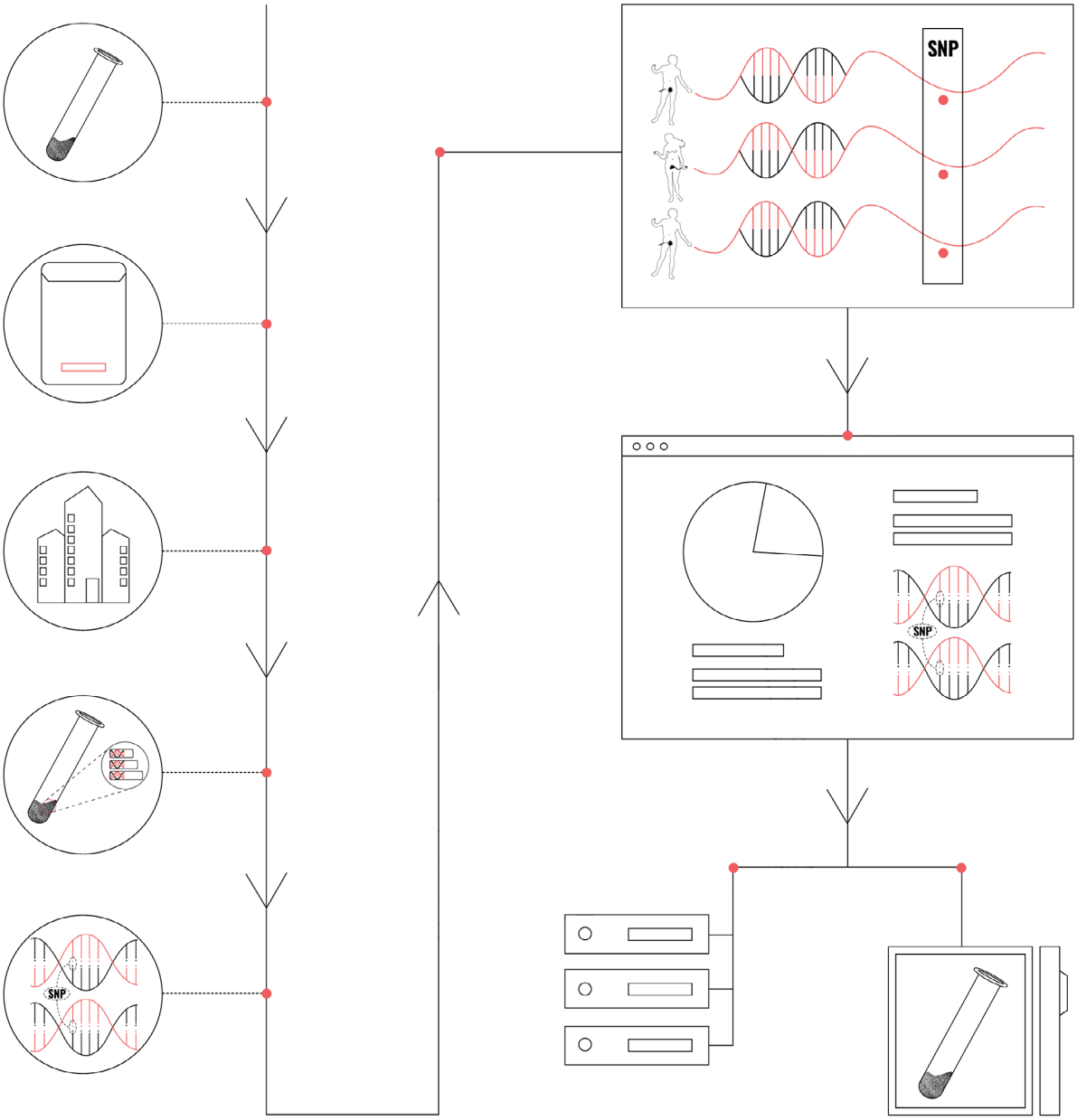
SO, HOW DOES IT WORK?

A customer spits in the tube and sends their saliva to a genetic testing company. When the company gets it, they first have to extract the DNA from saliva samples. The cell debris, along with proteins and all the material that is not DNA is being removed. Then, the copies of DNA are made using a technique known as Polymerase Chain Reaction (PCR), after which they are broken into shorter chunks.

These chunks are then fed into a machine called a genotyping array. These arrays work kind of like a coin sorter, but for SNPs. They will reveal which versions of SNPs you've inherited, and at what location in the genome.

After this process, the data obtained are compared with what is commonly referred to as a reference panel. A reference panel is made up of a selected number of DNA samples, obtained from previous customers who have taken the test and from publicly available DNA databases. The results that companies provide you highly depend on the size of their database – the bigger the database, the more accurate are the results. For example, AncestryDNA uses a database that contains more than 16,000 reference DNA samples from 43 regions around the world. About 12,000 of these samples come from Ancestry users who opt in and allow the company to use their DNA for research purposes, while the remaining reference samples come from public databases such as the 1000 Genomes Project.

However, there are cases when, for example, twins that are genetically basically identical, get clearly different results or that the same person, using different kits from different companies, or even from the same company in a period of time, obtain different results.



HOW DOES THIS ACTUALLY HAPPEN?

The bigger the database is in size, the results are more accurate. However, testing does not rely on database size only. A very important aspect to be considered is where the data comes from. For example, in the case of Ancestry.com, almost 75 percent of AncestryDNA's ethnic regions skew toward European descent, so at present it is very difficult to obtain detailed estimates of ethnicity from other regions. Along the same line, there are fewer reference samples from both consumers and scientific research available in regions of Asia and Africa, thus the accurate estimates for genetic heritage in those locations remain more prone to error.

Anyhow, not to be forgotten that even though genetic ancestry tests deliver precise percentages about our heritage, the reports are best thought of as estimates, based on imperfect data.

WHAT DO THESE COMPANIES DO WITH THE DATA AND SAMPLES?

All of the companies seem to have an explained protocol of what they do with generated data and DNA samples. For example, Ancestry.com states that once a laboratory partner has given DNA results, the DNA and saliva are stored for future testing. Moreover, further testing might be done in case the customer has signed an informed consent. In case that the customer does not want their DNA sample stored, they can contact the company and request the destruction of the sample. If customers previously agreed to research, Ancestry DNA's website claims it will not be able to remove genetic information from active or completed research projects. However, the company will not use genetic information after a requested deletion for any new research projects.

In a similar way, 23andMe ships customer's saliva samples to a third-party laboratory. Customers can choose whether to agree or not with storing their samples of saliva or DNA extracted from it.

Ancestry DNA:

- ***Service providers*:** Partner companies will have some of customers' information in their systems. Those companies include laboratory partners, DNA test shipping providers, payment processors, cloud services, etc.
- According to Ancestry DNA's website, "partners are subject to contractual obligations governing data security and confidentiality consistent with this Privacy Statement and applicable laws."
- ***Research Partners*:** When customers consent to genetic research, DNA information is shared with research partners. Check here for a list of Ancestry DNA's research partners.
- ***Legal obligations***: Ancestry DNA may share personal information if the company believes it is reasonably necessary to comply with valid legal process (e.g., subpoenas, warrants), enforce or apply Terms and Conditions, protect the security or integrity of the Services, or Protect the rights, property, or safety, of the company, its employees or users. Check here to see the number of law enforcement requests the company received.
- ***If acquired***: "If Ancestry or its businesses are acquired or transferred (including in connection with bankruptcy or similar proceedings), we will share your Personal Information with the acquiring or receiving entity. The promises in this Privacy Statement will continue to apply to your Personal Information that is transferred to the new entity."

23andMe:

- ***Service providers*:** While 23andMe explained it directly conducts the majority of data processing activities required to provide services, it works with some third-party companies which include order fulfillment and shipping, genotyping lab, cloud storage, marketing, etc.
- ***Commonly owned entities:*** "We may share some or all of your Personal Information with other companies under common ownership or control of 23andMe, which may include our subsidiaries, our corporate parent, or any other subsidiaries owned by our corporate parent..."
- ***Legal obligations:*** Under certain circumstances, 23andMe may share customers' personal information pursuant to laws, regulations, subpoenas, warrants, or orders.
- 23andMe will preserve and disclose any and all information to law enforcement agencies or others if required to do so by law or in the good faith belief that such preservation or disclosure is reasonably necessary..." Check here to see the number of law enforcement requests the company received.
- ***Business transactions:*** If the company goes through a merger, sale, or acquisition, personal information may be transferred. Customers' information would be held to the promises made in any pre-existing Privacy Statement, according to the company's website.

WHAT DOES THE LAW SAY?

Laws on personal data protection that apply to one's case might vary depending on various circumstances. For example, if the customer is in the EU, and data is processed by the company incorporated in the EU, the answer is simple - GDPR will apply. By the same logic, USA laws apply to companies incorporated in the USA. That is the easy part. The difficulties arise within the USA firstly because there is no USA federal data protection law, but each US state has its own legal regime. In addition, the GDPR rules will apply also to the USA company, if such a company processes data of EU citizens, under certain circumstances (so called "extraterritorial" jurisdiction rule).

But why is it relevant to determine which law applies to companies that provide DNA services? Because that relevant set of legal rules contains the question what are the obligations for the company, and what are the rights of their consumers / clients in relation to personal data processing. If you got the service as a State of Washington customer, your rights toward the company may differ from Californian or German customers. If you want to delete your data, you need to know based on which paragraph of which article of legal act you are entitled to make such a request, and the company has to comply. That is why with the mess in the applicability of different laws between the US and the EU, or any other country, might become a legal nightmare. If you need a lawyer to resolve it it could cost you much more than, what seemed at the time, an exciting self inquiry journey.

Good news is that there is an easier way out from the legal labyrinth. In fact, there are two. One comes from the fact that data protection laws and data subject rights are becoming increasingly harmonized over the western world. GDPR and California Consumer Privacy Act (CCPA) standards (applicable in California, directly applicable to e.g. Ancestry) are more or less the same. The other one is contained in the privacy statements and privacy policies of these companies, that are publicly available to consumers, and are legally binding for them. The bad news is that a customer who is a lay person might in any case need some legal assistance if they want to retrieve or delete their DNA, or any other personal data, gathered by these companies. That fun self inquiry took only spitting in the tube to kick off. But how long or complicated would be a procedure for data deletion if and when you don't find it so amusing anymore.

WHAT ARE THE INTERNAL PROCEDURES AND RULES OF THE DATA CONTROLLER?

If we take the privacy statement of Ancestry as an example, we can see at the outset it is a pretty detailed one. It aims to be reading-friendly and clear. However, when it comes to getting the answers to concrete questions about what they do with consumers data, there are visible holes.

Let's take a step back to see what types of responses we should be able to find there. In general, GDPR and CCPA (that are both expressly referred to by Ancestry in their privacy statement) regulate explicitly which types of information are to be provided to clients. More less - you can ask them any question in relation to processing of your data - where are they located, for which purposes are they used, on what legal ground is such usage happening, who has access to data, how long are they kept/stored and how are data protected, etc.

In addition, they must notify the clients (in data protection legal language we can call them data subjects) what are their rights under the applicable laws - including the famous right to be forgotten. Finally, it should be clear to whom a complaint can be made if certain rights, in the opinion of the client, are not respected in accordance with the law.

On the face of it, Ancestry provided all this information in their privacy statement. They did explain that after the customer gives them a saliva sample to get info about their ancestors this will lead to a number of future usages of the customer's various personal information for a variety of purposes. To begin with - there is a registration for their service and creation of a user account. They have an interface where a customer can see their results and a user-friendly manner to communicate with their service - including to request data deletion. But at a closer look it seems a bit more complicated to delete all they have on and know about a customer.

In concrete, according to the Ancestry privacy statement, here is what happens to customer's data after they spit in the tube.

ANCESTRY TAKES BIOLOGICAL SAMPLES

the saliva that is used to extract DNA. The saliva itself is not much of an interest before this step is taken, since there are no other procedures to be done with it in order to reveal who the customer is. However, once DNA is extracted - Ancestry has and knows customers' pure and whole identity.

However, according to the Ancestry privacy statement, neither saliva nor DNA extracted from it are customer's personal information - they are just biological samples. Strictly legally speaking, it seems they are trying to say that such biological samples are therefore not protected by data data protection laws.

While for saliva itself there could be arguments it does not constitute a personal data or information - in case there is no equipment necessary to extract DNA (which is not applicable to Ancestry, because they keep the saliva and can use it to extract DNA any time) or by relying on legal definitions that differentiate biological sample and data itself (such is definition of

"genetic data" in GDPR) - it is ambiguous based on what logic DNA is not considered the most personal of all personal data. Whatever logic Ancestry used for this differentiation is not explained in their privacy statement. Therefore, the legal status of biological samples kept by Ancestry remains unclear.

FROM BIOLOGICAL SAMPLES THEY TAKE GENETIC INFORMATION

according to Ancestry definitions, genetic information consist of: (i) DNA Data that are actually machine-readable biometric data converted from DNA which was previously extracted from saliva and (ii) any information derived from it such as ethnicity estimates, communities, traits, and relative matches. According to the privacy statement, DNA data is processed to provide the clients an ethnicity estimate and communities suggestions, as well as to provide other information about a customer such as degree of connection to genetic relatives in their database and any genetic markers associated with certain biological, physiological, or behavioral traits, e.g. hair thickness and eye color or traits associated with your health and wellness.

Since genetic information is considered personal information under Ancestry definitions, they should be protected by data protection laws also according to their privacy statement.

FROM BIOLOGICAL SAMPLES THEY TAKE GENETIC INFORMATION

according to Ancestry definitions, genetic information consist of: (i) DNA Data that are actually machine-readable biometric data converted from DNA which was previously extracted from saliva and (ii) any information derived from it such as ethnicity estimates, communities, traits, and relative matches. According to the privacy statement, DNA data is processed to provide the clients an ethnicity estimate and communities suggestions, as well as to provide other information about a customer such as degree of connection to genetic relatives in their database and any genetic markers associated with certain biological, physiological, or behavioral traits, e.g. hair thickness and eye color or traits associated with your health and wellness.

Since genetic information is considered personal information under Ancestry definitions, they should be protected by data protection laws also according to their privacy statement.

IN THEIR INTERACTION WITH THE CONSUMER, THEY MAY COLLECT MANY OTHER DATA

In addition to genetic information, other data that fall into this category are the account information, or user profile (if you have it), payment info etc.

However, a big problem with Ancestry's privacy statement is that it does not clearly delineate which data from this "other" category fall into their own definition of personal data. Namely, their definition of "Personal Information" is open ended and it reads: "[Personal Information] is information that can identify you directly or indirectly, as further described in [Section 3], such as your name, email or address, or information that could reasonably be linked back to you". In the further text, however, it is elaborated which particular information they have on their customers. Based on their source, these data are divided in the following three categories:

SECTION 3: "WHAT PERSONAL INFORMATION DOES ANCESTRY COLLECT FROM YOU?"

These are info about your account, user profile, your communication with other users while you are using the service, credit card and payment information, as well as DNA Kit Activation Information and Genetic Information.

SECTION 4: "WHAT INFORMATION DOES ANCESTRY COLLECT THROUGH YOUR USE OF THE SERVICES?"

These are computer and mobile device information, information from cookies and similar technologies, information shared through social media features and information from your use of the services.

SECTION 5: "INFORMATION WE COLLECT FROM OTHER SOURCES"

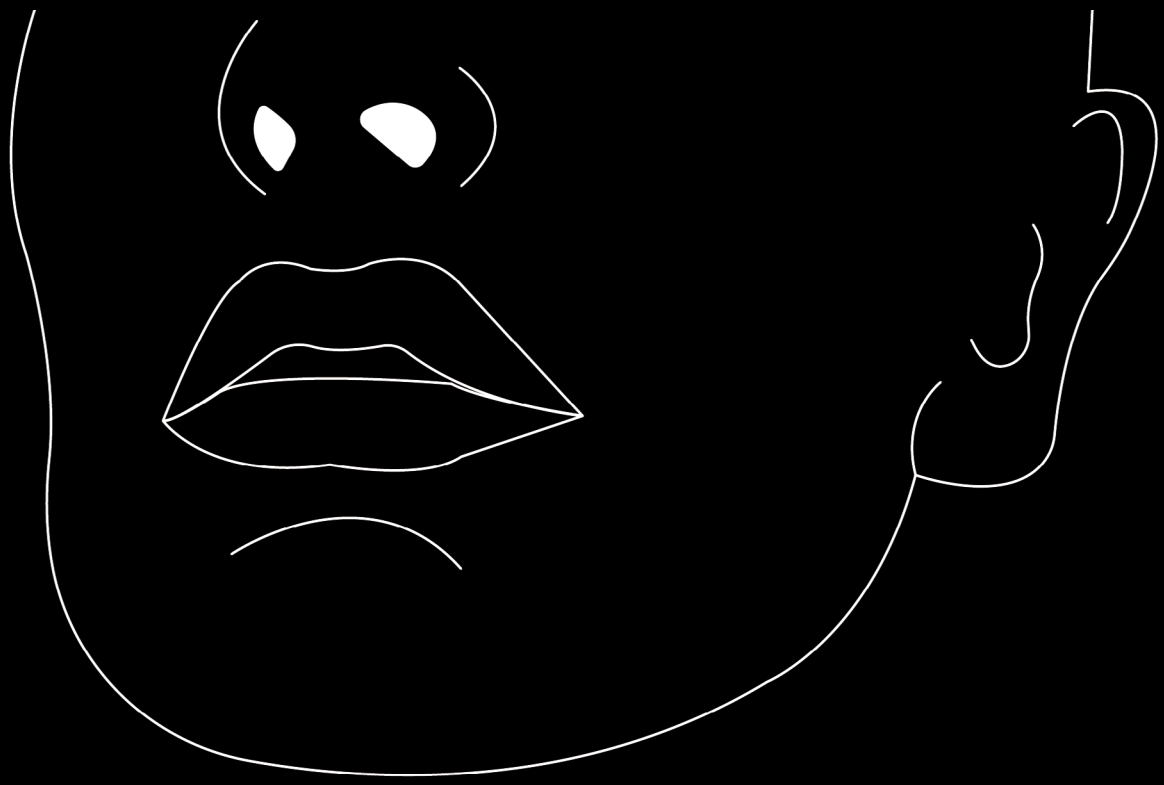
These are Information from public and historical records and information from third parties.

Since the Sections 4 and 5 do not say that the information listed therein are personal - could this also suggest that they are out of such definition and are not subject to laws and data subject requests for exercise of their legal rights? It should certainly not be the case, as it is contrary to both GDPR and CCPA definitions. But for some reason, Ancestry leaves in doubt whether they were just sloppy when they wrote their privacy statement, or if there is some legal intention behind it.

This is not purely an issue of legal gymnastics and perfectionism. Clarity and application of these definitions has a direct effect on the issue of which data are protected by privacy laws and which kind of legal rights can a customer exercise against the company.

If legal practitioners are having the trouble to understand what is Ancestry's position based on their privacy statement, it seems unreasonable to expect that an average customer can know what to do if they want to be completely forgotten by the company that has their DNA, in addition to bunch of other personal data collected along the way.

CAN ANCESTRY EVER FORGET YOU?



The main statement with respect to the question who holds the ultimate control of the information that resulted from giving a saliva sample to Ancestry should be the one at the beginning of their privacy statement which reads "You **always maintain ownership of your DNA and DNA Data** [...] you can manage and delete it as described in this Privacy Statement." With respect to other personal data the answer should also be short and simple. In the privacy statement, under the headline "**10. How Can I Delete My Personal Information?**" - the first sentence reads "You can delete your Personal Information from Ancestry at any time."

Again, when you continue reading privacy statements the words "always" and "at any time" turn out not to be what they seem.

Let's say a customer sends a deletion request for all their personal data, including DNA and biological samples and it has been accepted. Can they put all this experience behind and no one will ever know about it and there will be no possibility to track them back in any manner based on their relationship with Ancestry? Not quite.

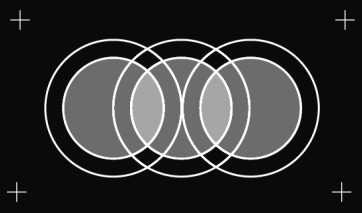
In the section 10 which elaborates on deletion procedure, a customer is also notified of the following: "Please note that there may be some latency in deleting your Personal Information from our backup systems after it has been deleted from our production, development, analytics, and research systems. Also, our laboratory partners may retain information they receive from us in order to comply with laws or regulations that may require them to do so, such as the Clinical Laboratory Improvements Amendments regulations administered by the U.S. Food and Drug Administration. Ancestry may also retain certain information as reasonably necessary to comply with our legal obligations (including law enforcement requests), resolve disputes, maintain security, prevent fraud and abuse, as well as to comply with tax, payment industry, securities, and clinical regulatory compliance requirements."

Of course, it would be Ancestry's discretion to determine what constitutes a "reasonably necessary" period to keep any customer's DNA for all these various and open-ended purposes. So, any customer that didn't spend some time to read the privacy statement in detail before deciding to "know their world from the inside" (as Ancestry nicely put it) by purchasing a DNA kit, might be unpleasantly surprised if they change their mind about sharing their inner world with a powerful multinational company (and any of their "trusted" partners).


DELETE.

To this day I am not sure why I decided to spit in the tube. The situation in terms of ancestry in my family is quite simple. We originate from the Balkans, a small, yet quite turbulent area in the southeast of Europe, so maybe I was expecting to discover something that might have happened during the Crusades, Ottoman Empire or any other time in history when big migrations and wars took place.

I discovered nothing exotic about my ancestry, so I decided that I wanted to delete both the digital and physical sample of my DNA Ancestry had on me, but that is another story.



2021