# New ways to plagiarise

Plagiarism of the future



UKÄ Rättssäkerhetskonferensen 2018 Jonas Lundqvist, Transcript (Google translate)

I would like to start by expressing my thanks to the Swedish Higher Education Authority who invited me to speak here today. Very exciting! For those of whom I have not had the pleasure of meeting earlier, I have spent a little over 18 years preventing, and if that failed - uncover - pupils and students attempts to take shortcuts. Without having seen the attendance list, I think that during that time I have had the privilege of visiting almost all the institutions represented here today at least thrice, either to discuss Plagiarism Prevention with your teachers or How to avoid plagiarism with your students. And a bunch of other topics. I said "until recently" because this Spring I chose, as the last remaining of the band's original lineup, to leave the antiplagiarism service which we started as five friends in 1999/2000. So I'm not here to talk about URKUND specifically – or at all, actually, other than to share some of my experiences from that time. 18 years of experience that makes it possible for me to be here today talking about what I'm here to talk about with relatively good confidence: as the title says "New Ways to Plagiarise" or "Plagiarism of the Future". I shall begin with the opposite, though, to give a little perspective. If you do not know the history, you are destined to repeat it.

## [Slide] (attr. George Santayana)

# [Slide] Ancient plagiarism

I will not go that far back in time. During the actual ancient times, plagiarism was often synonymous with "learning something" so we can skip quite a bit there. And I'm not really that old.

When I started with plagiarism prevention around the turn of the century, computers were still relatively new and difficult to get into. For grownups.

The casings the computers were in were called "hard drives".

- The CD tray was mistaken for a mug holder. More than once.
- Windows was called Word
- Computers never made a mistake

and the machines were fitted with a 5 second delay on the power button because it was common to shut down the computer in outright panic when making a mistake in a program – because of fear something had been irreparably broken.

*Young people* had a significantly better idea of how computers worked and what they could do and what the Internet really was. Some in the forefront of technology even realized that there were places on the Internet where you could get finished assignments and submit them - instead of writing them yourself.

And THAT was a much more effective method than using big sister's old book report

And THAT was a much more effective method than using big sisters old book report [Slide]

or retrieve a few years old essay from the student association's dusty archive

or get one from "buy and sell" in youth magazines like Rolling Stone - if you read it - something few teachers or professors did. And write it again in your own handwriting, or on a typewriter, or on a computer if you had one. And maybe adjust some parts that did not fit the circumstances. Phrases like "from my perspective as a woman" might not work so well when it's Erik who is submitting it...

#### [Slide]

As long as the recycling actually demanded some measure of labour from the would-be plagiarist, it was actually quite difficult for a teacher to detect when something was not right or, at least, to prove it, when the professional's senses and gut feeling sent up warning flags.

An original in a desk drawer or in a student association's cellar is not so easy to find. On the other hand, plagiarism was not that super-duper common, because it required manual effort. **[Slide]** 

When the process was digitised - and commercialised, it became much more common, and there were hushed whispers among students everywhere, about how easy it was to get homework done.

I attended a hearing with then Minister of Education Ingegerd Wärnersson at Rosenbad (the Swedish House of Parliament) in 2002 and there was this class from a building and construction high school there who gave a very telling giggle when the Minister asked them if THEY used to download finished homework of the Internet. It was no wonder they did.

There was not only Mimers Brunn, an essay site intended to inspire creative writing that was very popular with students, but also - if I remember correctly, some 35 self-confessed cheat sites in Sweden where you could download finished schoolwork and it was very difficult and cumbersome for a teacher to search those webpages to find a possible original version. Add to that, for space-saving reasons, the files were also compressed.

Studienet.dk, at that time a site with more dubious direction than it currently has, earned millions from selling school - according to (the business paper) Dagens Industri. Google was steeped in "cheat sites" and "paper mills" for some time before they, in 2008 I think it was, decided not to accept that kind of advertising or include them in search results anymore.

A good decision. It was a relief from our side not having to handle them anymore and explain why we did not get so many hits in their material - it was because they cost a pretty penny a pop and antiplagiarism services received no discounts - strangely enough.

With digitisation, however, it was also possible to develop TECHNICAL COUNTER MEASSURES and that's what we did. It was initially a very simple tool, not particularly advanced at all. A search engine we set to search in collected material, but in time it became increasingly complex. A tool that saved a lot of time for librarians and teachers, and helped to bridge some of the gap between students and teachers' IT skills.

Computer confidence and skill eventually increased among teachers - IT training campaigns, tax-free home PCs and the like, and in many ways, people caught up. The subculture that was Internet became everyday stuff for all.

[Slide]

Having an antiplagiarism system in an educational organisation became increasingly popular; even to the point where schools just claimed that they did to get the preventive, deterrent effect, free of charge. It was not only dissuasive, it was then and is now a great way to demonstrate to students that such behaviour is unacceptable, from both an academic and moral standpoint.

#### [Slide]

But, it wasn't acceptable before the computer made its debut either, and of course the students were aware of that already. Regardless of how many "I did not know I wasn't allowed to do that" one heard.

The reasons why a student then and now chooses to take a shortcut vary greatly, but they are almost always of a deep personal nature. There are circumstances like

Lack of time - as a result of every conceivable circumstance A knowledge deficit in the subject at hand and / or writing techniques Competition for opportunity or other desirable things

#### and more

Usually in combination with a moral compass that you may be able to ignore, at least temporarily, though it may sting a bit.

The decision is weighed against a result. A goal.

It may be praise from home, just enough points, a final grade, opportunity to move on to the next course, continued student grants and loans. A job, a career, a successful life.

Every single goal one that requires some amount character to deliberately abstain. The importance of these goals also varies over the world and socioeconomic factors play a significant role in many places where the need to succeed during ones education is critical.



Cheating in example to fairly common in the Indian state of Bihar, but new image, have emerged which show just how jarse-scale and biatant the practice is

These reasons and circumstances do not magically disappear because the possibilities of control became and becomes better and better. Or, if you as an educational organization make it clear that the solutions the students see to their plight, and sometimes choose, are reprehensible.

In situations and lifestyles where the goal is much more important than the perceived damage of taking a shortcut, a spiral begins. "It actually only affects me and I'm going to read up on this later!"

#### [Slide]

With consequences for getting caught established, and risks considered, pupils and students began to think about how to circumvent the controls.

#### [Slide]

https://www.flashback.org/t627620

(a talk-about-everything Internet forum in Swedish)

I'm pretty sure they started talking about us on Flashback earlier than 2008, but they may have cleared out some old posts lately.

On Flashback, people discussed, in-depth, methods of getting through undetected.

Like diligently change words to synonyms

putting the plagiarized text within quotes

Write the text in an image and paste it into the document or simply translate text from another language. And then clean it a little. Less now, or even at all. Translation services have become considerably better over the years

Teachers - in turn - eventually began to think that some still got through the controls and became more vigilant and began to raise questions. And we did our part by developing countermeasures against known tricks and evolving our algorithms and expanding the searchable material we had.

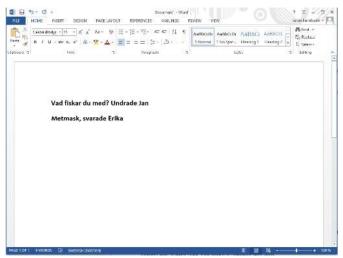
... And students and students began to think about how NOW to get around the controls

#### [Slide]

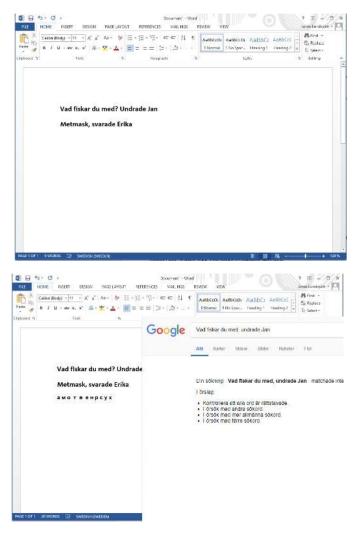
Like replacing letters with homoglyphs from other alphabets

or find characters that destroy text or can crash a check.

#### A simple text



(What are you fishing with? Asked Jan Earth worms, Erika replied)



Which can easily be given clues to by a search engine can become completely invisible to search engines with a bit of "search and replace". (Western characters replaced with Cyrillic look-alikes)

Such tricks, how creative they may be, are of course discovered sooner or later, and the world's antiplagiarism systems and tools patch them to the best of their ability as soon as they have the opportunity. Perhaps with a certain delay sometimes.

#### [Slide]

One loophole patched, students, or people who are NOT students, come up with a new way, which in turn is patched.

Sometimes, higher education institutions complement this development work with a ban on the act of making it difficult for control so that only an attempt to circumvent is damning enough in itself.

New methods to bypass different antiplagiarism systems spread quickly. And not just via Flashback but also via more internationally established information channels such as, Quora

#### https://www.quora.com/ls-there-any-way-to-cheat-the-plagiarism-software

#### Is there any way to cheat the plagiarism software?

Two ways: Method 1-(Usual method) STEP 1 - Open the word doc for which you want to check plagiarism. STEP 2- Replace the words with its synonyms (replace at least two words in a sentence) STEP 3- Replace unsolicited (say) with its synonym by right clicking it and then: Similarly replace maximum words to bypass plagiarism. Method 2-(easy and less time consuming) Basically, we will be replacing the blank spaces with the hidden alpha-numeric value whose font will be white in color to by pass plagiarism. STEP 1- Open the word doc for which you want to check plagiarism. Now go to Replace option: STEP 2- Find and Replace option will pop-up STEP 3- Go to More option: STEP 4- Now in the find what- type " ^w " (regular expression for blank spaces)

and in the replace with- type " i "(or any of the alpha-numeric value)

STEP 5- then go to format >> font (as done below)

STEP 6- Replace font color to White

STEP 7- Select Replace All option

Other powerful way is to use Blank symbol (from Character Map) instead of alphanumeric value (it may be not plagiarize 100% but is more efficient and less likely to be caught mainly by paid software). You may further need to use Method 1 to get 100% unique content.

To find Blank symbol, follow steps bellow:

Now check for plagiarism using any online available tool, it will be 100% unique. Hope it helped.

Where contributors usually are willing to answer any kind of question

#### [Slide]

#### Today, tomorrow, or at least, in the vicinity of now

When the antiplagiarism tools get so good that you can't nick text from the Internet or even from a book and then compile it with what you managed to produce yourself, what do you do if the need still exists?

Again check if there are any new ingenious ways to get around the controls, I suppose.

There are several digital services today that you can use to change text and with varying quality retain the meaning of its content and an understandable language.

Spinbot Article Spinning, Text Rewriting, Content Creation Tool.	Pricing	About	FAQs	API	Contact
Enter Text To Rewrite:					_

. . .

Characters Remaining (Max 10,000): Ignore Any:	10000	
Spin Capitalized Words: Remove Captcha and Ads Supercharge Your Workflow N	<u>ow!</u>	
Jag är inte en robot	reCAPTCHA Sekretess - Villkor	
Go! Rewritten Text:		

Such as Spinbot. Spinbot, however, isn't particularly good, I think, but there are others

Submit whatever you while maintaining the		hange it to something differen	Quill it
	Character Limit: 110 / 400	Log In / Sign Up for more charact	Prioritize by Changes Grammar V
		C ReQuill Copy Tex	t

Like Quillbot

C SEO Magnifier	Home Contact US Blog Priking Log In Sign Up
Best Paraphrasing Tool To Paraphrase Online	· · · ·
Advertisment	Advertisment
Online Paraphrasing Tool	
Daplash Action Processing Paraphrasing Segretions V Unique Action	
Paste (Crif = V) your article below then click Submit to watch this article rewriter do it's thing	121
Select Language:	
English	· •
Paraphraic It	

#### SEO Magnifier

HOME	OME BLOG ADVERTISE CONTACTUS BUY TEXTLINK AD HERE							电 LOGIN 🛔 R	EGISTER
PREF	POSTSEO	Plagiarism Checker	Word Counter	DA PA Checker	Reverse Image Search	Premium Plans		Search Seo Tools	٩
			Free	Online	e Article	Rewriter	1		
							_		
f v					ADVERTISE HERE				
St	ep 1 ste Duplicated Art	icle	Step 2 Processing		Step 3 Re-writing		Step 4 Done (Unique	e article)	
0	Get Google Chron	ne Extension						∰ Report a	a bug
									4
0	OR / Belect File doc / doc/ / pdf / bit								
	ADVERTISEMENT								
					ADVERTISE HERE				
				1	Rewrite Article				

And PrepostSEO

These are services where you paste text and then let an algorithm rewrite it into something that has the same content but that is said - not in YOUR OWN words - but in other words.

When you think about it, one might wonder if this is technology that might not have existed to the same extent if there were no antiplagiarism tools. The last service even has its own antiplagiarism tool built-in, which may be symptomatic.

[Slide]

By the way, while we are using online services, why not let the algorithms write the text altogether?

For example, Articoolo

ant (	articcolo services	<ul> <li>HOW IT WORKS PRICE LIST</li> </ul>	LOG I	sign UP
	Create	unique textual co	ontent in a flash wr articles	
			RITE ARTICLE	
	LANGUAGE Write me ar ENGLISH	n article about (describe it in 2-5 words)	1	
AGMACRIA E INO FERRET				
	an in fair and a start of the s			Q
1111	and and Pallin and a start	P/P	116	A

or any other provider of services that creates texts on a topic that you decide on, based of real sources.

Quora	Home Home	Answer	Notifications	Q Search Quora
Aziz Be		eter and Blogg	ger and <mark>F</mark> ather of Tv	vo
-	que Content (	Generator On	line! 🖉	
2. Unique A	Article Creato	r 🖉		
3. Article G	enerator Pro	ď		
4. SEO Arti	cle Generator	Unique Art	icles for your Blog	S 🖉
	Article Creato - Purchase 🖻		ilder, Article Spini	ner, Article Creator
article ge Rewriter money. I need of c	enerator i can is a great arti t still takes tin	suggest spin cle spinner w me to create a ld links to yo	riter Ever ⊮ if you o rewriter .idit is All which will save you a good readable ar ur money site, the at this tool	l in one, Spin 1 a lot of time and ticle. If you are in
		sometimes us	ing Spin writer 🗗 🛛	but i do my
oroofreading 5.2k Views - Viev				

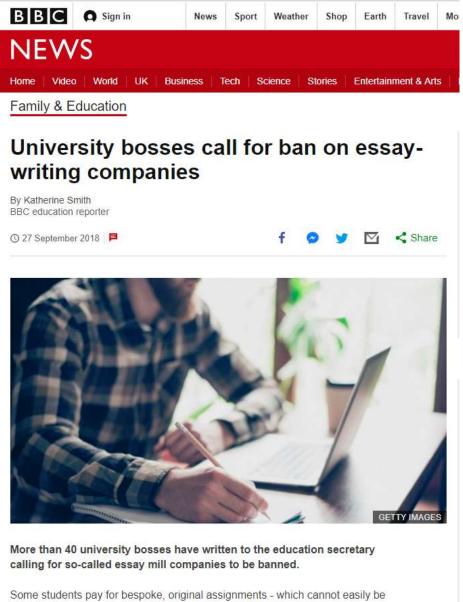
This is a technology that is currently best suited for shorter texts. Currently. But many snippets can very well become an entire essay.

It is also of course possible to employ human resources by purchasing a unique document, complete with source references in the right place and literature list and everything - by someone else. Contract Cheating. Ghostwriting services.

And then we're back in big sister's desk drawer again.

## [Slide]

Apparently, a problem of such magnitude in England that legislation is being discussed.



detected by anti-plagiarism software.

But the letter says these essay-writing services should be targeted rather than

Once more unto the breach, comes programs for author recognition, sometimes included in antiplagiarism tools. Algorithms working through texts proven to be by the student, and matching

the pattern with the latest work - or the opposite, matches a newly submitted document to known ghost writers, if you know any - with varying degrees of accuracy to the statistical probability of it being the same author.

However, in the world of school and academics, such technology can collide with such things as active feedback - by a mentor, companion, parent or by all means using software, joint efforts during a group assignment, and with personal development. You change your style the more you learn. But author recognition is a relatively new technology that will obviously evolve over time and there are many EdTech companies who are investing properly on it right now.

#### [Slide]

Perhaps the technology may be made to take such factors into consideration, so it can be used to reveal that the person submitting a text has not written it - without at the same time delivering false positive warnings and causing potentially grave problems for honest students

#### [Slide]

I used to be very clear when I lectured to students, because I always got the question, that you should not think so much about how to get around an antiplagiarism system as they are now. As a student, you should instead ask yourself which technology may be available in 5, 10 or 20 years when their texts, perhaps, may be checked again.

And what will happen in the future? I thought it could be a nice delimiter between the past and the present and the future with some dramatic music so I'll use 3 of my 45 minutes for that.

## https://youtu.be/6k-7yxNF3UM

## The Future

Certainly, Shakespeare is debated as to who actually wrote what, and authors recognition programs are of course used to try to investigate this, and a lot of Macbeth is said to come from other works, but that's not why he's with us here. Shakespeare was included because he is featured in conjuncture with technology that he could never have imagined. Or what us five aspiring young EdTech entrepreneurs could imagine 18 years ago either. Artificial Intelligence.

First, it should be said that it is a term whose meaning seems to be all the more loosely defined these days and it is currently used for much much that actually has their very own names and abbreviations, but let's call it AI for now. It will sort itself out eventually.

I will return to Shakespeare in a moment, but first few words about the music you heard. It's not impossible that it had pleased Shakespeare and it fits quite well with any of his plays, but it's not written by any composer contemporary with him or even of any modernist like Mozart or Beethoven - or even a human being.

It is written by an AI to be the lead motif of a computer game - Pixelfield "Battle Royal" and personally I think it managed quite well.

#### [Slide]

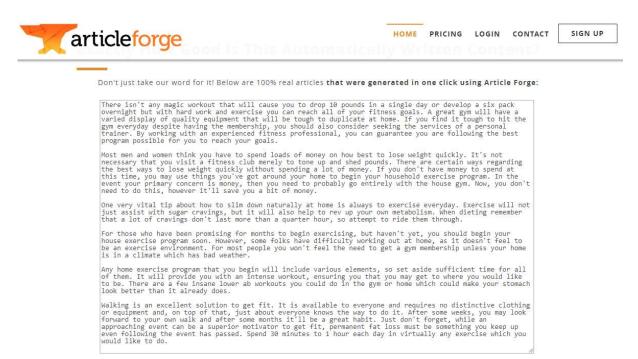
If you follow TED talk, or want to give it a once-over, there is a very interesting talk by Pierre Barreau, who has built AIVA which is behind this music, explaining how it works. In short, AIVA analyzes existing classical music and draws lessons on what works well and what works less well and then creates something new and beautiful from it. AIVA made the music you heard shortly at the beginning of this presentation as well.

The technology behind services like Articoolo, as you saw previously, uses language templates and facts it retrieves from real sources for writing their texts, but it cannot be used to create something new. Just condense and rephrase what is already established.

With AI you can make something new. IBM's has a AI called "Chef Watson" and it is used to create new exciting recipes - which, in fact, some say is not only edible but delicate. Recipes and computers seem to go hand in hand through history.

On a more serious level, US law firms use AI to compose new arguments and interpretations of the law that are more successful than lawyers can accomplish themselves. Other companies use AI to compile and draw conclusions from medical data and patient descriptions of side effects.

There is no big leap from "dumb" services like Articoolo, or like this Articleforge,



that are already writing articles and maybe even simplistic school work to "smart" systems that can achieve actual academic content.

Oops! It appears that already happened

"At CTI we believe that education is important to all of us and we are passionately working to transform it. Using our proprietary AI, we can create educational content at a price point that can't be matched.

Our AI systems can create with a staff of 20 what it would take a large publisher with a thousand employees to replicate. We are breaking down the cost barriers and have proven that AI can produce better content in less time at a fraction of the cost. Talk about leveling the playing field!

We provide solutions for secondary education, higher education, continuing education and corporate training. Have you own content? No problem, we can repackage and repurpose for you."

http://contenttechnologiesinc.com/

Let's listen to their CEO

http://contenttechnologiesinc.com/ CEO Video

Rather impressive if the sales pitch is to be believed

But author recognition...? Should not such technology be able to recognize that it is someone other than the alleged author who wrote a text? Does it matter that the author is an AI? [Slide]

Just like an AI can be given a set of texts can draw conclusions from, it can be provided with the author's profile and be instructed to use that style for the final product. It has already been done with written text and that's where Shakespeare, in the end, makes his entrance.

IBM Research Australia, University of Toronto and University of Melbourne have let an AI study a few thousand existing sonnets and then applied the style Shakespeare used - and accomplished poetry that may well be mistaken for being the Great Poet

"With joyous gambols gay and still array, no longer when he 'twas, while in his day at first to pass in all delightful ways around him, charming, and of all his days."

No, I'm not going to attempt to recite that.

A Recurring Neural Network, RNN, a kind of AI, have after reading 5000 pages of Game of Thrones, wrote a first draft of the next book in the series - and with varying success sought to imitate how George R. Martin writes. As a matter of fact, it became more comical than serious but, of course, it is just the beginning.

Well. Ok. Can't *another* AI be used to recognize if an AI has written the work, even with that type of templates and customizations? Yes. No. Perhaps. How reliable the probability calculation would be, is a bit difficult to say. Who has the best AI?



Image from <a href="https://www.youtube.com/watch?v=mpw\_FB2QrjQ">https://www.youtube.com/watch?v=mpw\_FB2QrjQ</a>

#### [Slide]

So what conclusions can you draw from development of late?

# [Slide]

• AI is just a fad

No, I jest. AI is definitely the future. It has the capacity to transform products and services that today can seem advanced and complex to bucket and spade-level.

[Slide]

• AI will permeate our entire society

It will not happen tomorrow. And not the day after tomorrow. But "maketh no mistake" it's not far away. It will happen step by step but fast, relative earlier development, as it is an area where huge sums of money are invested on research. Just as AI will revolutionize all other areas, it will also happen within the higher education.

Here are some news hot of the presses, from October 15th. MIT is building a faculty for AI research. Not primarily for the computer sciences, but for other subjects. Such as Biology, Chemistry, History and Language Studies. For a billion dollars.

When AI can write academic work containing new conclusions, one might ponder how this method of examination and academic merit will work at all, but that is a matter for another forum.

## [Slide]

• Artificial Intelligence will sooner or later be able to imitate a human writer to perfection

## [Slide]

• Artificial Intelligence will sooner or later be able to produce texts with academic quality

Al will also play a role in learning itself, which will probably help to break the arms race of tricks and fix that has been the norm up to now, by addressing the underlying problems – and, dear audience, THAT is fantastic.

"Already, there are intelligent tutoring systems such as Carnegie Learning that use data to provide feedback and work with students directly.

These tools are designed to support teacher and tutor approaches to student difficulties but soon will be more advanced and capable of providing specific details for students as well."

Al offers an opportunity to tap into the adaptive learning processes already being featured in assessment software, learning games and digital textbooks to individualize learning. Tools that can highlight and emphasize key areas where students are suffering allow teachers to focus on facilitating the learning process and offer the one-on-one support that students at all levels need."

https://www.forbes.com/sites/theyec/2017/12/27/how-ai-impactseducation/#24f85190792e

# [Slide]

• Artificial Intelligence will, quite soon, be able to help students with the specific problems they, personally, are struggling with during their studies

In the slightly shorter perspective and if classical dissertation will be relevant in the future and if the underlying factors that drive a pupil or a student to take an shortcut by plagiarism still remain

relatively unchanged, the attempts to get circumvent antiplagiarism services with various tricks will continue with the same regularity as before.

Someone somewhere realizes that 18 text layers in a PDF containing 17 white Lorem Ipsum text makes the percentages of the reports drop to such low numbers that further review is not prioritized - the antiplagiarism systems counters with automated OCR of the text that actually appears or warns for an abnormally large amount of characters in relation to page numbers.

Someone else realizes that a certain combination of characters in the right place can get a review to go wrong. The antiplagiarism systems starts automatically looking for them and removing them. And so on, unt so weiter.

# [Slide]

• Attempts to find technical loopholes will continue and they will be patched eventually

Online services that rewrite text or create new text based on content from actual sources develop at the same rate as other technologies. They are completely legal, usually free, easy to use and since all texts are quite unique, they are also very difficult to find as there are no origin to match against

In spite of achievements in semantic analysis, the burden of proof using author recognition to examine that type of shortcuts will be difficult to handle.

## [Slide]

• Public digital tools for creating and rewriting texts will become better and better and will be more often

[Slide]

• Public digital tools will eventually take on AI

Although some have been deterred, and some even become better writers of fear of getting caught by mistake, students and students in general have not stopped plagiarizing. Instead, with the help of technology, they have become better plagiarists. With that in mind, academic institutions will spend more time solving the underlying causes of students taking shortcuts, trying to break the negative spiral.

# [Slide]

• New models of examination and methods for how to design assignments aimed at maximizing the chance to get genuine material in return will see increased interest

#### [Slide]

• Students will themselves seek out and use legitimate digital support systems, such as Grammarly, to a greater extent than today.

#### [Slide]

• Administrative supervisor system, with a structured approach to essay writing, such as SciPro, a Swedish system, will become more common

# [Slide]

If I really peer into the crystal ball here at the end, one should not underestimate the possibility that there will eventually be a perfect lie detector readily available and that the question

"When you wrote your essay, did you do something that you knew was breaking the rules?"

...might become really hard.

#### [Slide]

That was what I had. Many thanks again to the Swedish Higher Education Authority for the opportunity to be here today!

