The function I want to present in this blog post is present the problem that is addressed by the research. Because remember, either the authors explain the value of their readers and thus gain the attention of readers, or they don't and consequently lose readers' attention.

The fact of the matter is, background information just for the sake of background information is not research. Background information just like that does not lead to understanding, and the aim of research is to increase understanding. Therefore, you must show the value of what you have researched.

Authors create value by choosing a problem that (a) readers recognize as being in need addressing or that (b) readers want to see addressed. Authors then establish the connection between that value and their own research by performing the function of present the problem. In this way, authors reveal at least one aspect by which the problem can approached. Essentially, the authors pave a way toward solving the problem, and this way passes directly through their own research findings.

The function present the problem is really one subfunction to the overarching function of create value for the research. Therefore, present the problem must appear at two absolutely crucial locations in every paper:

> *(Location No.1) In one half of the Title, preferably the second half.*
>
> *(Location No.2) In the opening sentences of the Abstract, preferably the first three.*

To illustrate how these locations are utilized for the function present the problem, I have used the dblp to obtain search results for "machine unlearning" from the year 2020. I have kept papers only from prestigious conferences either in machine learning or in security and privacy. Of the remaining papers, I have kept again only those which are well-cited. (Citation counts have been drawn from Google Scholar and normalized for the year 2023.)

The Titles

Here you see the ten Titles, with the second halves highlighted in yellow — I have defined half *either as half the words or as one more than half the words or as that portion which follows a colon.*

Title 01 (Bourtoule et al. S&P 2021) *Machine Unlearning*

Title 02 (Chen et al. CCS 2021) *When Machine Unlearning Jeopardizes Privacy*

Title 03 (Sekhari et al. NeurIPS 2021) *Remember What You Want to Forget: Algorithms for Machine Unlearning*

Title 04 (Brophy et al. ICML 2021) *Machine Unlearning for Random Forests*

Title 05 (Gupta et al. NeurIPS 2021) *Adaptive Machine Unlearning*

Title 06 (Thudi et al. USENIX 2022) *On the Necessity of Auditable Algorithmic Definitions for Machine Unlearning*

Title 07 (Warnecke et al. NDSS 2023) *Machine Unlearning of Features and Labels*

Title 08 (Thudi et al. EuroS&P 2022) *Unrolling SGD: Understanding Factors Influencing Machine Unlearning*

Title 09 (Marchant et al. AAAI 2022) *Hard to Forget: Poisoning Attacks on Certified Machine Unlearning*

Title 10 (Nguyen et al. AsiaCCS 2022) *Markov Chain Monte Carlo-Based Machine Unlearning: Unlearning What Needs to Be Forgotten*

As always, we begin at an exception.

Title 01 is hardly worth even dividing, and it certainly doesn't present the problem, unless just naming the topic can be called present the problem. Nonetheless, this is the best paper (by citation count) in the whole batch, by far! It is my opinion that Bourtoule et al. do in fact present the problem by simply naming the topic; in other words, machine unlearning itself is in need of advancement, and Bourtoule et al. propose in their work how that advancement could be made.

Similarly, Title 05 is not a good illustration of Title halving. Still, here it might be said that Gupta et al. have enlisted the front half of their Title to imply that, as machine unlearning is currently applied or understood, the technique fails at adaption.

Now I make my point. Look at Titles 02 and 04 and 07 and 08 and 09 and 10 — all of these put the second half of the Title to work at presenting the problem. For example, Titles 04 and 07 make plain the case that either a method or an application of machine unlearning is not performing well or only poorly understood in the research.

That leaves just Titles 03 and 06. The authors of these have simply decided to utilize not the second half but the first half of the Title in order to present the problem. This is a perfectly acceptable way to write. However, I would make just this one remark: Notice how more authors choose to utilize the second half of the Title. The authors of 03 and 06 are bucking the trend, and bucking the trend has generally one of two outcomes: either you draw extra attention to yourself or you simply get misunderstood. That's the choice, so choose wisely.

The Abstracts

Here are the relevant portions of the Abstracts — and you will see, 3 is indeed the average number of sentences needed by these authors to present the problem.

Abstract 01 (Bourtoule et al. S&P 2021) `Once users have shared their data online, it is generally difficult for them to revoke access and ask for the data to be deleted. Machine learning (ML) exacerbates this problem because any model trained with said data may have memorized it, putting users at risk of a successful privacy attack exposing their information. Yet, having models unlearn is notoriously difficult.`

Abstract 02 (Chen et al. CCS 2021) `The right to be forgotten states that a data owner has the right to erase their data from an entity storing it. In the context of machine learning (ML), the right to be forgotten requires an ML model owner to remove the data owner's data from the training set used to build the ML model, a process known as machine unlearning. While originally designed to protect the privacy of the data owner, we argue that machine unlearning may leave some imprint of the data in the ML model and thus create unintended privacy risks.`

Abstract 03 (Sekhari et al. NeurIPS 2021) We study the problem of unlearning datapoints from a learnt model. The learner first receives a dataset S drawn i.i.d. from an unknown distribution, and outputs a model w that performs well on unseen samples from the same distribution. However, at some point in the future, any training datapoint $z \in S$ can request to be unlearned, thus prompting the learner to modify its output model while still ensuring the same accuracy guarantees.

Abstract 04 (Brophy et al. ICML 2021) Responding to user data deletion requests, removing noisy examples, or deleting corrupted training data are just a few reasons for wanting to delete instances from a machine learning (ML) model. However, efficiently removing this data from an ML model is generally difficult.

Abstract 05 (Gupta et al. NeurIPS 2021) Data deletion algorithms aim to remove the influence of deleted data points from trained models at a cheaper computational cost than fully retraining those models. However, for sequences of deletions, most prior work in the non-convex setting gives valid guarantees only for sequences that are chosen independently of the models that are published. If people choose to delete their data as a function of the published models (because they don't like what the models reveal about them, for example), then the update sequence is adaptive.

Abstract 06 (Thudi et al. USENIX 2022) Machine unlearning, i.e. having a model forget about some of its training data, has become increasingly more important as privacy legislation promotes variants of the right-to-be-forgotten. In the context of deep learning, approaches for machine unlearning are broadly categorized into two classes: exact unlearning methods, where an entity has formally removed the data point's impact on the model by retraining the model from scratch, and approximate unlearning, where an entity approximates the model parameters one would obtain by exact unlearning to save on compute costs.

Abstract 07 (Warnecke et al. NDSS 2023) Removing information from a machine learning model is a non-trivial task that requires to partially revert the training process. This task is unavoidable when sensitive data, such as credit card numbers or passwords, accidentally enter the model and need to be removed afterwards. Recently, different concepts for machine unlearning have been proposed to address this problem. While these approaches are effective in removing individual data points, they do not scale to scenarios where larger groups of features and labels need to be reverted.

Abstract 08 (Thudi et al. EuroS&P 2022) `Machine unlearning is the process through which a deployed machine learning model is made to forget about some of its training data points. While naively retraining the model from scratch is an option, it is almost always associated with large computational overheads for deep learning models. Thus, several approaches to approximately unlearn have been proposed along with corresponding metrics that formalize what it means for a model to forget about a data point.`

Abstract 09 (Marchant et al. AAAI 2022) `The right to erasure requires removal of a user's information from data held by organizations, with rigorous interpretations extending to downstream products such as learned models. Retraining from scratch with the particular user's data omitted fully removes its influence on the resulting model, but comes with a high computational cost. Machine "unlearning" mitigates the cost incurred by full retraining: instead, models are updated incrementally, possibly only requiring retraining when approximation errors accumulate. Rapid progress has been made towards privacy guarantees on the indistinguishability of unlearned and retrained models, but current formalisms do not place practical bounds on computation.`

Abstract 10 (Nguyen et al. AsiaCCS 2022) `As the use of machine learning (ML) models is becoming increasingly popular in many real-world applications, there are practical challenges that need to be addressed for model maintenance. One such challenge is to 'undo' the effect of a specific subset of dataset used for training a model. This specific subset may contain malicious or adversarial data injected by an attacker, which affects the model performance. Another reason may be the need for a service provider to remove data pertaining to a specific user to respect the user's privacy. In both cases, the problem is to 'unlearn' a specific subset of the training data from a trained model without incurring the costly procedure of retraining the whole model from scratch.`

Notice how for this research focus of machine unlearning the authors really present the much the same problem. Plain and simple, the problem is this: Data must be removed from models, but the models must continue to operate; therefore, model training needs adjusting, since a complete retrain is inhibitive. Now I know, that skips some details. For example, Abstract 09 presents the more specific problem of the practicality of current formalization. Nonetheless, the problem data-must-be-removed-we-need-to-adjust-the-training — *that* pretty much captures the whole thing, doesn't it?

So, once again, the point to note is the majority choice. You choose a problem in a research focus in unison with the important papers on that research focus. Because if you choose otherwise, readers may well not get that your paper is about the same focus as the papers that they know.

You can, of course, present a new problem in the focus. However, to do so, you will need to be certain that your results are that novel and your interpretation that convincing to warrant a new direction in the research focus. Because that will be, essentially, what you are proposing: a new direction for an established focus. That sort of a work comes along only once in a while. Is yours really such a work?