How Artificial Intelligence can change the pharmaceutical landscape

Program & Abstracts

IBSA FOUNDATION for scientific research

How Artificial Intelligence can change the pharmaceutical landscape

XVI Forum

October 9th 2019, Lugano

Program & Abstract

IBSA FOUNDATION for scientific research

Percorsi Editoriali

Chairman

Andrea Danani	Istituto Dalle Molle di studi sull'Intelligenza Artificiale (IDSIA), Lugano, Switzerland; Università della Svizzera italiana (USI), Lugano, Switzerland; Scuola universitaria professionale della Svizzera italiana (SUPSI), Lugano, Switzerland

Speakers

Alessandro Curioni	IBM Research lab, Zurich, Switzerland; Swiss Academy of Engineering Sciences, Zurich, Switzerland
Gianni De Fabritiis	ICREA, Barcelona, Spain; Universitat Pompeu Fabra, Barcelona, Spain; Acellera, Barcelona, Spain
Sean Ekins	Collaborations Pharmaceuticals, Inc., Raleigh, United States of America; UNC Eshelman School of Pharmacy, University of North Carolina at Chapel Hill, Chapel Hill, United States of America
Ed J. Griffen	MedChemica Ltd., Biohub, Alderley Park, Macclesfield, Cheshire, United Kingdom
Jürgen Schmidhuber	NNAISENSE, Lugano, Switzerland; Swiss AI Lab Istituto Dalle Molle di studi sull'Intelligenza Artificiale (IDSIA), Manno, Switzerland; Università della Svizzera italiana (USI), Lugano, Switzerland; Scuola universitaria professionale della Svizzera italiana (SUPSI), Lugano, Switzerland
Alex Zhavoronkov	Insilico Medicine, Hong Kong; Buck Institute for Research on Aging, Novato, United States of America; Biogerontology Research Foundation, London, United Kingdom

Scientific Committee

Andrea Danani	Istituto Dalle Molle di studi sull'Intelligenza Artificiale (IDSIA), Lugano, Switzerland; Università della Svizzera italiana (USI), Lugano, Switzerland; Scuola universitaria professionale della Svizzera italiana (SUPSI), Lugano, Switzerland
Silvia Misiti	IBSA Foundation for scientific research, Lugano, Switzerland

Program

13.30 **Welcome of IBSA Foundation**
Silvia Misiti

Introduction
Andrea Danani

Chairman: Andrea Danani

14.00 **True Artificial Intelligence for learning robots**
Jürgen Schmidhuber

14.25 **Emerging challenges for Artificial Intelligence in medicinal chemistry**
Ed J. Griffen

15.00 **Can we machine learn chemistry and drug discovery?**
Gianni De Fabritiis

15.40 *Coffee break*

16.00 **Machine learning for rare and neglected disease drug discovery**
Sean Ekins

16.30 **New small molecule design pipelines utilizing generative models and reinforcement learning accelerate drug discovery**
Alex Zhavoronkov

17.05 **Artificial Intelligence and the future of discovery**
Alessandro Curioni

17.30 **Closing remarks**

with the exceptional participation of artist **Martin O.**

Index

9 **Presentation**
Silvia Misiti

11 **Introduction**
Andrea Danani

13 **AI and healthcare in a superfluid market**
Andrea Danani

25 **True artificial intelligence for learning robots**
Jürgen Schmidhuber

26 **Emerging challenges for Artificial Intelligence in medicinal chemistry**
Ed J. Griffen

28 **Can we machine learn chemistry and drug discovery?**
Gianni De Fabritiis

29 **Machine learning for rare and neglected disease drug discovery**
Sean Ekins

30 **New small molecule design pipelines utilizing generative models and reinforcement learning accelerate drug discovery**
Alex Zhavoronkov

33 **Artificial Intelligence and the future of discovery**
Alessandro Curioni

37 **Biographies**

Presentation

Silvia Misiti
IBSA Foundation for scientific research

The mission of the IBSA Foundation is to promote a science culture and to serve as a meeting point between the scientific world and the general public.

In order to achieve this goal, IBSA Foundation focuses on a range of activities, the most important of which is the organization of international Forums: these are one-day meetings that cover different and evolving aspects of new frontiers of life-science-related subject areas and bring a global network of scientists together to discuss the latest pre-publication research in their fields.

IBSA Forums – always organized in collaboration with the academic world, institutions, and leading research centres – represent a chance for speakers to compare and contrast ideas, as well as to exchange information and ideas on forward-looking topics and new developments in scientific research.

They also represent an opportunity for students, experts, and all participants to discuss issues, share insights and learn: a process of exchange that provides valuable input for further research and drives advancements in knowledge.

In each Forum prominent experts from the international scientific community focus on a significant topic in the field of biology or medicine that has a great impact on people's health and quality of life, describing state-of-art and future challenges.

The "IBSA Foundation Papers" series brings together the proceedings of Forums which the Foundation has organized since its establishment in 2012.

All Papers can be downloaded for free in PDF format from ibsafoundation.org and are also available in print version.

Introduction

Andrea Danani
Istituto Dalle Molle di studi sull'Intelligenza Artificiale (IDSIA), Lugano, Switzerland; Università della Svizzera italiana (USI), Lugano, Switzerland; Scuola universitaria professionale della Svizzera italiana (SUPSI), Lugano, Switzerland

Artificial intelligence (AI) is gaining more and more importance in the pharmaceutical sector, deeply transforming the drug discovery process. There are many potential benefits of applying AI techniques to improve the development of new molecules and the identification of new targets, cutting R&D costs and time.

In drug discovery, AI helps in predicting the efficacy and safety of molecules. It gives researchers a much broader chemical pallet for the selection of the best molecules for drug testing and delivery. In this context, drug repurposing is another important area where AI can have a substantial impact. With the extensive amount of clinical and pharmaceutical data available to date, AI algorithms find suitable drugs that can be repurposed for an alternative use in medicine.

Moreover, with the help of AI, it becomes easier to run clinical tests, diagnose diseases and provide the most effective treatment for a particular disease. As it can interpret test results, AI can also look through various sources including publications to correctly diagnose critical ailments.

New approaches based on AI in the drug discovery process and in the repositioning of old molecules will be presented during this particular Forum. We hope to provide an overview of the sectors where AI might have an impact on the pharmaceutical pipeline and play a crucial role in the pharmaceutical world in the coming years.

AI and healthcare in a superfluid market

Andrea Danani
Istituto Dalle Molle di studi sull'Intelligenza Artificiale (IDSIA), Lugano, Switzerland; Università della Svizzera italiana (USI), Lugano, Switzerland; Scuola universitaria professionale della Svizzera italiana (SUPSI), Lugano, Switzerland

Healthcare in a superfluid market

For centuries, frictionless markets have matched buyers and sellers. However, as trade and technologies became increasingly complex, companies started to be formed. In this new environment, market friction created companies that were trading structures between businesses and countries, and also employment structures between companies and employees. As a result, traditional markets were often viscous, slow, expensive and opaque, with the main player being the company, not the consumer.

As a result of the Industry 4.0 revolution and the emergence of increasingly sophisticated technologies powered by mobile, additional opportunities appeared to empower consumers. Traditional friction that provided companies with security started to evaporate and markets that were once viscous became fluid or even superfluid, leading to a closer connection between demand and fulfilment, and disrupting traditional rigid trading structures by democratizing data, shrinking distance and eliminating intermediaries.

During this time of great disruption, a company accumulates business intelligence by dealing with challenging assumptions, and not from doing tomorrow what you did yesterday. Disruptive factors do not always appear as critical inflection points. Similarly, innovation can come from many different angles and is often more a mindset than a technology or a product. Therefore, the ability to take advantage of disruptive factors is fundamental for a company's success. Like in a living organism, advance receptors must detect and understand upcoming environmental changes and having the right organizational design allows this early detection to happen so that a company

can be ready. This is why a traditional structure may not be able to pick up these new signals.

The healthcare market does not yet seem to have reached this superfluid state and an important question that life sciences companies must address is how to seize the upside of this disruption, harnessing the power of this innovation.

Indeed, the overall experience of healthcare remains disjointed, as people are shuffled between multiple caregivers, and regulations and data incompatibilities make it difficult to easily share the information necessary to optimize care. The ubiquity of data and analytics continues to blur the traditional boundaries between therapeutics, medical technologies, consumer devices and information technology.

Against this backdrop, platforms are interfaces that allow existing and new stakeholders to connect and share data and insights quickly, while simultaneously deriving benefits from the act of participating. These structures are the backbone of the future sharing economy and have demonstrated enormous value across different industries. Researchers at the MIT School of Management revealed that in 2013, 14 of the top 30 brands by market capitalization belonged to platform companies. Amazon, for instance, has transformed the way we shop; Airbnb has transformed leisure travel; Uber and Lyft have transformed the transportation industry. What do these platforms have in common? All are convenient to use and focused on the user experience. All simplify access to the best products and services and create an opportunity for networks defined by different resources and capabilities to emerge. All drive profitability by eliminating intermediaries. Perhaps most importantly, all combine data in new ways to reveal new business opportunities based on needs that were previously unrecognized. How might platforms unlock similar value in the field of healthcare?

In healthcare, platforms become a means to connect highly disparate stakeholders - consumers, physicians, payers, policymakers and product makers - and combine their capabilities. This process should allow the consumer to become the central focus of companies. The development of technology to capture or connect data will help to find personalized solutions for health-related issues linked in particular to diagnosis, treatment and changes in behavior. To address this point, an essential requirement is the aggregation and storage of data so that it can be easily shared in real time (• Figure 1).

Health data is isolated in databases dispersed across a wide range of organizations (see • Figure 2). Because of the way these databases are structured, it is difficult to interpret and use the data to improve health results.

• Figure 1. The building blocks to create platforms of care

Health data will be aggregated and structured to create usable insights at the therapeutic-specific level and the supra-platform level.

Point solutions

EVOLVING RAPIDLY

Companies develop solutions or services tied to specific products that operate in isolation.

Therapeutic-specific platforms of care

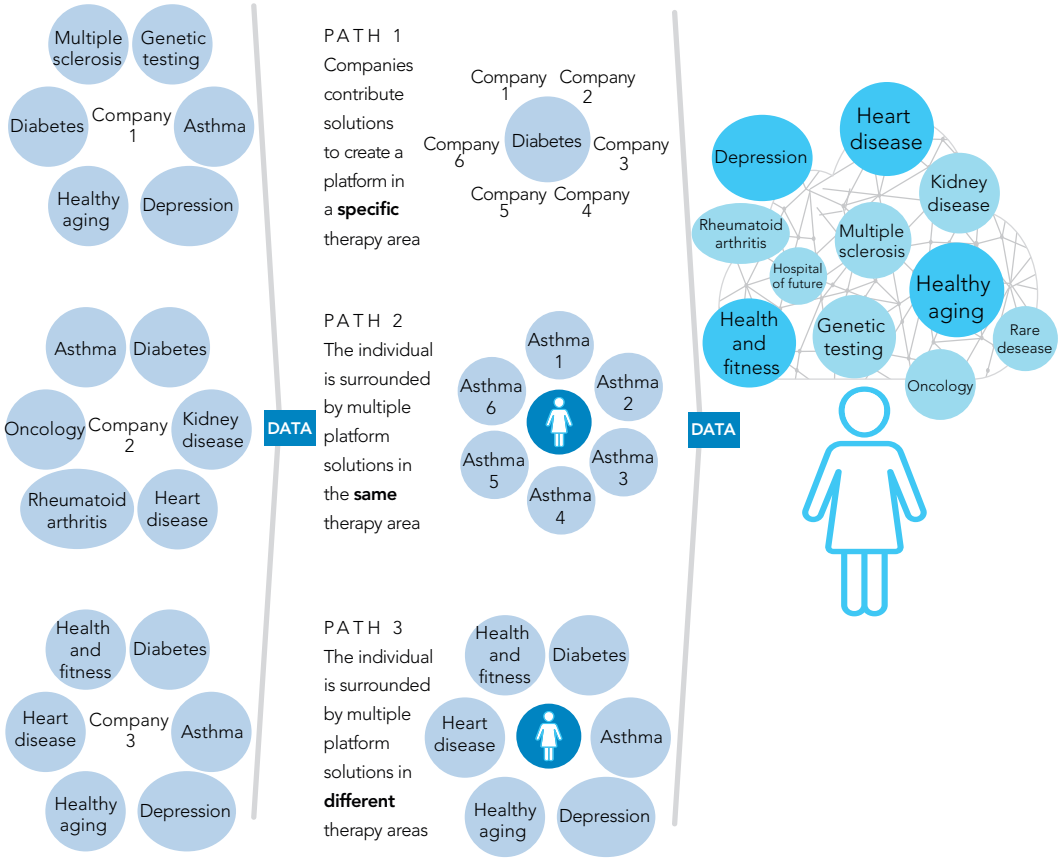
EMERGING

Products, tools and applications from different developers are integrated making it possible to connect, combine and share a variety of data.
Three different paths to the formation of therapeutic specific platforms of care are possible.

Supra-platform aggregation

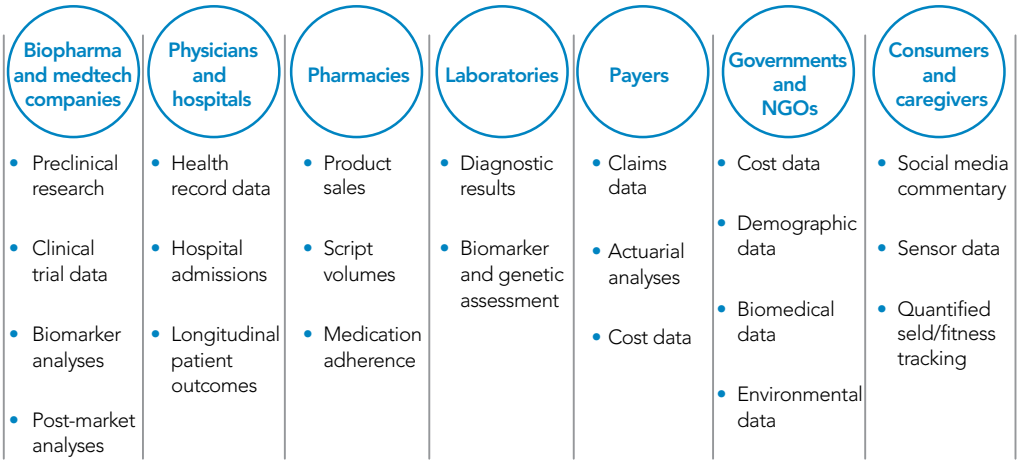
FUTURE STATE

Organizations – seamlessly and without friction – connect, combine and share data to establish integrated platforms of care that provide a convenient and easy-to-use interface with the patient/consumer at the center.



Source: Life Sciences 4.0: Securing value through data-driven platforms, Progression, EY, 2018.

• Figure 2. Health data are siloed across many different organizations



Source: Life Sciences 4.0: Securing value through data-driven platforms, Progression, EY, 2018.

Even when companies have significant reservoirs of health information within their own organizations, a significant proportion of it goes unused. Clearly, given their data expertise, technology companies are best placed to drive the integration required at both the platform of care and supra-platform levels. However, life science companies should participate in the formation of medically-relevant platforms of care either aimed directly at consumers or developed in conjunction with physicians or payers. Indeed, for life science companies, developing products that make people want to stay involved in their care will be essential. Forging stronger links with their end-users, they can incorporate additional activities that lead to the disappearance of entrenched corporate relationships within the healthcare cycle.

Fortunately, public announcements indicate that a breakthrough may come sooner than expected. Recently, various organizations have joined forces to create new health services and products that are more convenient and customer-focused. By offering consumers a wider array of options in how, and where, they receive their medical care, and who has access to this information, these initiatives could redefine the care delivery experience.

Such adaptations are likely only the beginning as the convergence of AI, robotics and 3D printing creates opportunities for consumers and physicians to turn health data into actionable information - especially when linked together via emerging blockchain-enabled infrastructure platforms. As wearables become both more powerful and smaller in the future, they

will help nudge consumers to make better health and lifestyle decisions that ultimately result in improved outcomes. When this happens people will not orbit companies, but companies will start to orbit individuals.

The rise of deep learning

Artificial intelligence (AI) is at an exponential development stage. People are jumping into it, spin-offs and start-ups are mushrooming left and right, and everyone is looking for a business with a winning idea. The strong feeling is that all that is happening around AI will not be slowing down any time soon and there seems to be no going back.

One of the main reasons for the impetuous growth in AI in the last ten years has been the availability of so-called GPUs (graphics processing units) - the chips on graphics cards that were designed for graphic-intensive games. But, why are GPUs getting so much attention now? The answer lies in the rise of deep learning, an advanced machine-learning technique that is widely used in AI and cognitive computing. Deep-learning powers a number of different fields, including autonomous cars, cancer diagnosis, computer vision, speech recognition and many other cases of intelligent use.

Like most machine-learning algorithms, deep-learning relies on sophisticated mathematical and statistical calculations. Artificial neural networks (ANN), convolutional neural networks (CNN) and recurrent neural networks (RNN) are some of the modern implementations of deep-learning. These neural nets emulate the human brain with a close resemblance to neuroscience. Each type of neural net is aligned with a complex use case like classification, clustering and prediction. For example, image recognition and face recognition use CNN, while natural language processing relies on RNN. ANN, the simplest of all the neural networks is often used for predictions involving numerical data.

Irrespective of the type of neural network used, all deep-learning algorithms perform complex statistical calculations. From typical stock market data to the picture of a cat to a radiology report, everything is first decoded into a set of numbers. In the case of images, the first step is to turn them into grayscale and then to assign a number to each pixel depending on how light or dark it is. As we can imagine, a simple selfie shot from a mobile phone translates into a few million pixels, which in turn translates into a large matrix of numbers. During the training phase of deep-learning, these matrices of numbers are fed as input into the neural network along with the correct classification. For example, by training the neural network with thousands of cat images, we are going to get a model

that can easily recognize a cat visible in a photo. This training process is all about correlating multiple pixels (numbers) in order to find patterns of a cat image. The correlation involves multiplying millions of matrices with one another to achieve the right result. To increase the training speed, these operations need to be done in parallel.

Typical central processing units (CPU) are designed to tackle calculations in a sequential order, where each mathematical operation will have to wait for the previous to complete. CPU with multiple cores may marginally speed up the calculation by offloading the operations to each core, but they are prohibitively expensive, making them not as good for training neural networks. On the other hand, with a GPU, we have a processor with thousands of cores capable of performing millions of mathematical operations in parallel. This is why in this new era of AI, GPUs could find their place. There is a similarity between graphic rendering and deep-learning: both scenarios deal with a huge number of matrix multiplication operations per second. That's one of the reasons why laptops or desktops with high-end GPUs are preferred for deep-learning. Now, another class of GPU called GPGPU, which is a general-purpose GPU meant for parallelization is powering deep-learning experiments in advanced research and development. To put things into perspective, Nvidia's latest GPUs come with more than 5000 cores, while Intel's top end server CPUs can have a maximum of 28 cores. However, the rise of GPUs does not result in the end of CPUs. We still need these bulky processors to do the heavy lifting. The combination of CPUs and GPUs along with sufficient RAM makes it a perfect testbed for deep-learning.

AI and medicinal chemistry

But what is the situation in medicinal chemistry? Back in the late 1980s, the advent of high-powered computational facilities (at that time) heralded a new era in computational drug design. Here was a technology that might render largely obsolete the irrational and biased designs of the medicinal chemist and deliver that single molecule designed to bind specifically to any given drug target with exquisite potency and selectivity, fast-tracking the delivery of molecules through to clinical evaluation.

After three decades, the reality is somewhat different. Many successes have been certainly achieved: structure-guided drug design has accelerated the path to clinic for many therapeutic agents; analysis of drug target structure has expedited unprecedented levels of selectivity; the understanding of metabolic enzymes, at the atomic level, has allowed the prediction and

optimization of the molecular stabilities of emerging drug candidates; and QSAR modelling has enabled the application of accrued data to predict and prioritize the biochemical and pharmacokinetic properties of possible new compounds for synthesis. However, despite ever-increasing computational power and better force field models to describe interaction, the original dream of computational chemistry has yet to be realized and scientists now accept these limitations, work alongside them, using computational models in a realistic manner as an adjunct, rather than a panacea of prediction.

After these decades of incremental innovations in drug design, the tuning of AI applications in drug discovery has started to take root only in the last years. AI tools have started to show promising results and according to the Global Market Insights report, the drug discovery sector currently represents 35% of the AI applications market, where as expected, the application of deep-learning plays an important role.

The ultimate goals of applying AI and ML methods to challenges in drug discovery remain the same as they ever were: bringing the best drugs to the clinic to satisfy unmet medical needs. These new concepts and the associated AI technologies have the potential to improve the drug discovery process as well as making pharma R&D more efficient by automation. In early discovery, AI can support finding better drugs against well-characterized targets by analyzing and predicting druggable binding sites and exploitable interactions, as well as for target deconvolution/identification. Deep-learning algorithms can transform screening programs by being taught to design optimized compounds and increase hit rates using the knowledge encoded in large published and proprietary datasets. Further applications involve repositioning and redesign programs that search known sets of drugs to find pharmacotherapies for new indications. Drug repositioning is an effective strategy for finding new indications for existing drugs and is highly-efficient, low-cost and without risk. Compared to traditional drug development strategies, there are less steps in drug repositioning: compound identification, compound acquisition, development, and FDA post-market safety monitoring. Due to the fast growth of bioinformatics knowledge and due to the availability of large biological databases, drug repositioning decreases the time cost of the drug development process significantly. Drug repositioning appears to be an obvious candidate for deep-learning both because of the large amount of high-dimensional data available and the complexity of the question being asked. However, even if massive machine-learning algorithms have been introduced to improve the performance of drug repositioning in this regard, a mature and production-ready framework for such drug repositioning is currently in its early stages and needs further work in order to be effective.

Medicinal chemistry teams could gain multiple benefits from these tools. In drug design, AI technologies are perfectly suited for solving multidimensional optimization problems that are typically the case in drug discovery settings.

In synthetic chemistry, deep-learning can suggest viable alternatives on how to best synthesize designed compounds. The optimization of chemical reactions contributes to providing the best compound at the best time that may deliver lower attrition rates as compounds move through the pharma value chain. The efficiency of these processes is often enhanced by automated technologies in both synthesis and testing that provide a previously unseen amount of data feeding and improvement of AI-based design technologies.

Finally, AI technologies may have a significant impact on clinical research. Its advancement to the results of clinical trials makes the process faster, cutting costs, improving trial quality, and reducing trial times.

We conclude with the following important question: is deep-learning superior to other machine-learning methods? In other words, has deep-learning transformed the study of human disease? Although the answer is highly-dependent on the specific domain and problem being addressed, we believe that deep-learning has not yet realized its transformative potential or induced a strategic inflection point. Despite its dominance over competing machine-learning approaches in several biomedical and pharmaceutical areas, deep-learning is still far from having "solved" these problems. Deep-learning has a much more flexible architecture than other machine-learning approaches, and hence it is possible to create a neural network architecture that is tailor-made for a specific problem. However, a disadvantage is that deep-learning in general needs very large training sets. For this reason, we believe it is still too early to draw any conclusion.

In medicinal chemistry specifically, the results thus far indicate that deep-learning is very efficient for de novo molecular design and reaction predictions. For tasks with structured input descriptors, deep-learning seems to perform at least on-par with other methods. The most relevant example is bioactivity prediction, where deep-learning seems to achieve a better overall performance, but it is worth noting that other machine-learning methods are also improving. If different machine-learning methods achieve roughly the same accuracy, the limit of what can be achieved with a machine-learning model could depend on the experimental uncertainty of the size of the data and dataset rather than the specific algorithm used. Thus, in practice, the choice of method may depend on which method the modeler is most familiar with.

We are witnessing deep-learning models achieving human-level performance across a number of biomedical domains. However, machine-learning algorithms are also prone to mistakes that humans are much less likely to make, such as the misclassification of adversarial examples, a reminder that these algorithms do not understand the semantics of the objects presented. It may be impossible to guarantee that a model is not susceptible to adversarial examples, but work in this area is ongoing. Therefore, human experts and deep-learning algorithms still need to work together in order to address many of these challenges and achieve better results than either would individually.

Abstracts

True Artificial Intelligence for learning robots

Jürgen Schmidhuber
NNAISENSE, Lugano, Switzerland; Swiss AI Lab Istituto Dalle Molle di studi sull'Intelligenza Artificiale (IDSIA), Manno, Switzerland; Università della Svizzera italiana (USI), Lugano, Switzerland; Scuola universitaria professionale della Svizzera italiana (SUPSI), Lugano, Switzerland

Our deep learning artificial neural networks have won numerous contests in pattern recognition and machine learning by imitating human teachers. They are now used billions of times per day by the world's most valuable public companies, whose main profits are in marketing.

True AI, however, goes far beyond this – that's why for over 3 decades we have also been working on unsupervised AIs & robots that invent their own experiments and goals, to figure out how the world works, and to become increasingly general problem solvers.

I will discuss recent state-of-the-art examples. What will be the next wave of AI? In the not too distant future (not yet!) we will have "show-and-tell robotics" or "watch-and-learn robotics" or "see-and-do robotics:" quickly and economically teach a recurrent neural network to control a complex robot with many degrees of freedom to execute complex tasks, such as assembling a smartphone, solely by visual demonstration, and by talking to it, without touching or otherwise directly guiding the robot – a bit like we'd teach a kid.

Finally, I'll speculate how learning robots will transform every aspect of our civilization, and eventually colonize the universe, and make it intelligent.

Emerging challenges for Artificial Intelligence in medicinal chemistry

Ed J. Griffen
MedChemica Ltd., Biohub, Alderley Park, Macclesfield, Cheshire, United Kingdom

The application of artificial intelligence (AI) to improving medicinal chemistry is at a pivotal point. Significant advances have been made in accelerating data extraction, computer aided synthesis planning, compound property prediction, de novo design and integrating systems for automated drug discovery. The current major challenge is persuading chemists to use these systems.

AI approaches have demonstrated significant success in other domains, particularly in image recognition, recommender systems, advertising, navigation assistance and game playing. However, in all these cases the consequences of an incorrect recommendation are minor.

Often the lack of confidence in the system is implicit in that multiple recommendations are made simultaneously; multiple routes may be suggested or simply that it is just a game. When systems address problems where the consequences of a poor recommendation become higher, such as clinical decision support, or autonomous vehicle navigation systems must be auditable. Systems used by experienced professionals in a challenging scenario where the effect of a poor decision may have a career impact need to able to "explain their thinking" for the skilled user to trust them. These domains are now being addressed by what has been termed "explainable AI".

Medicinal chemists are highly skilled and operate in a domain where the risks of failure are very high. Considering the overall cost of synthesizing a compound and all the assays and experiments it might require to prove a compound is poor may be in the range $ 1,000-25,000. Obviously, this influences medicinal chemists' behavior. Over the last decade we have developed systems that are highly transparent and can be audited by the medicinal chemist.

We have focused on the common pharmacokinetic and toxicology problems in compound design as these are applicable across multiple target classes and disease indications. Matched molecular pair analysis (MMPA) is a technique that has been used informally to understand the SAR of sets of molecules for decades.

In its simplest form it is the identification of pairs of compounds connected by a chemical transformation, such as the replacement of an aromatic proton with an aromatic halide. As the size of the compound set (n) grows, the number of potential pairs increases with n^2. Rapidly it becomes almost impossible to perform the task without computational support. The local chemical environment needs to be captured as part of the transformation between pairs of compounds. These aspects represent significant algorithmic and engineering problems. We demonstrate how large rule databases can be built from ADMET datasets by the application of pair-finding algorithms and how this can be updated automatically to build an a "learning machine". The rules discovered can be used to appraise ideas for synthesis, and to perform de novo design suggesting new molecules for synthesis. This is a very transparent form of "explainable AI", because the underlying compound structures and measured data supporting the rules can be seen. Examples of successes in drug hunting projects will also be shown.

MMPA is a very "data hungry" approach and therefore we assessed other matching learning approaches to evaluate how transparent or "explainable" they are. Although popular, neural network architectures have a key weakness in their lack of explanation. We therefore assessed k-nearest neighbor (kNN) and Random Forest models which can be coupled with improved descriptor methods to show the user the molecules driving their predictions.

We believe these "explainable AI" systems represent a useful approach combine machine learning methods and human insight, delivering more rapid project progress and show examples of their use in modelling a range of cardiovascular and seizure related toxicities.

Can we machine learn chemistry and drug discovery?

Gianni De Fabritiis
ICREA, Barcelona, Spain; Universitat Pompeu Fabra, Barcelona, Spain;
Acellera, Barcelona, Spain

Machine learning in drug discovery is not new, as classical machine learning has been used for decades, however in recent years there has been great advances in artificial neural network applications and a new wave of methods for solving drug discovery problems.

In our laboratory, we are trying to work within this latest approach trying to machine learn chemistry and in so doing obtain levels of generalization beyond previous methods. There are essentially four aspects: ML predictors, generative models, ML potentials and active learning agents.

In our recent activity we demonstrated state-of-the-art in several predictors relevant for drug discovery, in particular binding affinity, pose, ADME, etc. by winning multiple blind challenges and also by tests with our industrial collaborators. While these methods might be state-of-the-art, they are not always based on deep learning methods which does provide some useful indications. I will also present one application of novel structure-based generative model, their current performance and limitations.

ML potentials will play a very important role in the near future, we have worked to use and develop ML potentials both at the quantum and at the macromolecular level. All these tools are available at playmolecule.org. PlayMolecule represents our personal view on a drug discovery gym environment and in this context, agent-based learning can have an important role. I'll show some examples of agents, including outside of the domain, like in the case of 3D environment represented by the Unity obstacle tower challenge.

Machine learning for rare and neglected disease drug discovery

Sean Ekins*
Collaborations Pharmaceuticals, Inc., Raleigh, United States of America;
UNC Eshelman School of Pharmacy, University of North Carolina at Chapel Hill, Chapel Hill, United States of America

In an effort to address rare and neglected diseases Collaborations Pharmaceuticals, Inc. was formed and to date we have obtained 10 NIH grants since 2016 for projects including Chagas disease, HIV, Ebola, tuberculosis, Batten disease and machine learning software development. In addition, we have collaborated with various groups to develop machine learning models for different diseases.

Our approach leverages a wide array of computational approaches (Bayesian, SVM, Random Forests, Deep Neural Networks, etc.) to enable the selection of compounds for testing and to organize our efforts efficiently so we can work on many projects simultaneously.

To address more of the 7,000 rare diseases, we need technologies like this to build the scale that can handle orders of magnitude more diseases. To date we have applied our software called Assay Central to published studies on the estrogen receptor, Ebola virus, tuberculosis, HIV reverse transcriptase and numerous individual targets and ADME/Tox datasets such as CYP2B6.

Our company represents a prototype of a new approach to develop treatments that may be outside the current realm of big pharma and biotechs. The science as well as the challenges and opportunities of entrepreneurship will be highlighted.

* Sean Ekins[1], Kimberley M. Zorn[1], Thomas R. Lane[1], Daniel P. Russo[1,2], Daniel H. Foil[1], Ana Puhl Rubio[1], Jennifer Klein[1], Alex M. Clark[3] and Maggie A.Z. Hupcey[1]

[1] Collaborations Pharmaceuticals, Inc., Raleigh, United States of America; [2] The Rutgers Center for Computational and Integrative Biology, Camden United States of America; [3] Molecular Materials Informatics, Inc., Quebec, Canada.

New small molecule design pipelines utilizing generative models and reinforcement learning accelerate drug discovery

Alex Zhavoronkov
Insilico Medicine, Hong Kong; Buck Institute for Research on Aging, Novato, United States of America; Biogerontology Research Foundation, London, United Kingdom

Generative adversarial networks (GANs) and Reinforcement Learning (RL) techniques enable the generation of small molecules with the desired medicinal chemistry properties. In our previous work we demonstrated the theoretical applications of generative models for de novo molecular generation [1-4] and provided basic experimental validation [5]. Generative Adversarial Tensorial Reinforcement Learning (GENTRL) model for de-novo molecular design [6].

In GENTRL we combin reinforcement learning, variational inference and tensor decompositions, into an original generative, two-step machine learning algorithm: first, we infer an underlying manifold of a huge number of the reported compounds; second, we expand it to discover new scaffolds. We parameterize the structure of the learned manifold in the tensor train format to use *partially known* properties (*vide infra*).

By combining several approaches, the algorithm is quite robust: it avoids mode collapse, it uses partially labeled data, and it extrapolates the chemical space of molecules, thus suggesting novel valid structures which are not presented in the training libraries.

This model avoids the drawbacks of generative models reported previously: GANs can suffer from mode collapse, namely they fail to suggest entire chemical classes of compounds attractive for drug discovery. Autoencoder-based generative models, such as variational autoencoders (VAE) and adversarial autoencoders (AAE) avoid mode collapse by compressing the space of structures onto a simply organized latent distribution usually following a standard Gaussian distribution. Simple latent distributions facilitate ML model training but may not be optimal mappings for the

topology of the latent space for chemical structures.

The proposed prior distribution used in this paper flexibly parametrizes the latent space in as a high-dimensional lattice with an exponentially large number of multidimensional Gaussians in its nodes. This parametrization better ties the latent codes and properties, and works with missing values without explicitly imputing them (in contrast to other semi-supervised models). Previous reinforcement-learning-based approaches used reward functions that are less relevant for the generation of new structures because they used rather trivial rewards, for example the predicted n-octanol/water partition coefficient, traditional drug-likeness rules, and synthetic accessibility.

The current model uses self-organizing maps (SOMs) to estimate how chemical structures relate to reported small-molecule kinase inhibitors, incorporating temporal structure-based trends observed in intellectual property for drug discovery. The method employs the following features:

- *Trending SOM* is a Kohonen-based reward function that discriminates "novel" compounds from "old" compounds considering the application priority date of chemical structures disclosed in relevant patents. Compounds claimed during different time periods are located in distinct areas within the Kohonen map. Neurons abundantly populated with novel chemical entities have been used to positively reward the generative model, driving it closer to clusters which contain new molecules;
- *General Kinase SOM* is a Kohonen map that reliably distinguishes kinase inhibitors from other classes of molecules, guiding the ML engine towards kinase-biased structures. These two groups are uniformly spread along distinct areas of the map, providing a statistically significant separation. Structures predicted as kinase inhibitors with the cell coefficient over 1.3 were then subjected to the specific Kohonen-based classifier;
- *Specific Kinase SOM* was trained to isolate DDR1 inhibitors from the total pool of kinase-targeted molecules. We observed that DDR1 inhibitors were distributed within an ensemble of closely-related neurons at the periphery of the map. GENTRL prioritized the generated structures using the above combination of SOMs, thus at the final step placing more emphasis on nodes occupied by DDR1 inhibitors.

Here we demonstrate a case study where GENTRL is challenged to produce a small number of molecules with the specified properties in the shortest possible time for a known protein target, DDR1, implicated in fibrosis. All know publicly-available inhibitors of DRR1 were used for training. Within 21 days using GENTRL we produced a set of 6 molecules, 4 of which were

active against DDR1, 2 of these molecules were tested in metabolic assays and in mice confirming microsomal stability and favourable pharmacokinetic properties. In the pharmaceutical industry this process usually takes several years.

This study demonstrates that significant acceleration of drug discovery processes is possible using the recent advances in artificial intelligence. We later test one of these molecules in cancer patient-derived xenografts (PDX) in mice and demonstrated efficacy.

References

[1] Kadurin A, Aliper A, Kazennov A *et al. The cornucopia of meaningful leads: Applying deep adversarial autoencoders for new molecule development in oncology.* Oncotarget 2016;8(7):10883-90.

[2] Kadurin A, Nikolenko S, Khrabrov K *et al. druGAN: An advanced generative adversarial autoencoder model for de novo generation of new molecules with desired molecular properties in silico.* Mol Pharm 2017 Sep 5;14(9):3098-104.

[3] Putin E, Asadulaev A, Vanhaelen Q *et al. Adversarial threshold neural computer for molecular de novo design.* Mol Pharm 2018 Oct 1;15(10):4386-97.

[4] Putin E, Asadulaev A, Ivanenkov Y *et al. Reinforced adversarial neural computer for de novo molecular design.* J Chem Inf Model 2018 Jun 25;58(6):1194-204.

[5] Polykovskiy D, Zhebrak A, Vetrov D *et al. Entangled conditional adversarial autoencoder for de novo drug discovery.* Mol Pharm 2018 Oct 1;15(10):4398-405.

[6] Zhavoronkov A. *Deep learning enables rapid identification of novel potent DDR1 kinase inhibitors.* Nat Biotech 2019, in print.

Artificial Intelligence and the future of discovery

Alessandro Curioni
IBM Research lab, Zurich, Switzerland; Swiss Academy of Engineering Sciences, Zurich, Switzerland

About 80 percent of the world's data is tucked behind the firewalls of organizations. These incumbent organizations have extracted data in abundance from activities in both the online and physical domains. The data they have accumulated is proprietary. It's theirs to exploit and yet most organizations admit they fall far short of utilizing it. In other words, they have big data but too little insight and value.

The gap between those who use data to great effect and those who fail to do so is growing. Organizations that have adopted the platform business model, for example, have doubled down on data and typically achieved revenues eight times the average. These platform operators excel at using data and analytics holistically to inform their business strategy and also to hone continuous performance improvement. The data they possess is well integrated and fit for purpose.

For too many, the enterprise-wide integration of data remains more ambition than reality. Fewer than 4 in 10 organizations have integrated their data across the enterprise, or designed and deployed an enterprise-wide data architecture.

Moreover, the target keeps moving. Instead of catching up, organizations find themselves falling further behind. For example, organizations recognize that it's no longer sufficient to simply pour all their data into a data lake and expect everyone to go fishing. Instead, organizations must curate and enrich data so that it meets specific business needs, functions and workflows. A marketer, for example, needs a very different view of data than the deep data dive preferred by R&D.

As yet, fewer than half of organizations surveyed in our 2018 IBM IBV Global Artificial Intelligence (AI) Study have developed a systemized approach to data curation. To do so requires robust data governance, driven by business needs, and metadata. Metadata, the information about information, describes both content and context. It depicts the relationships among data, its source and history.

Absent robust metadata, it's difficult to integrate data and retrieve it for a specific purpose, or to deploy analytics and AI models. Agreed-upon governance practices are necessary to set the parameters, concur on "labels" and decide what metadata is needed. Machine learning technologies can then be trained to profile and classify data, automating metadata generation. "Cartridges" – pre-built sets of industry-specific data – are also a way to streamline curation of heterogenous data.

Unstructured and real-time data presents other challenges. Organizations train virtual agents, for example, by having them "listen" to recorded voice calls. Voice is converted to text and then a summation of the context of the calls, the answers to questions asked, and a prioritized list of dialogues is produced. The data, however, must be maintained and linked in its various states, including voice, text and summarized Q&A. This is because often when an organization completes one phase of a project, data analysts are finally smart enough to go back to the beginning and ask better questions.

If organizations don't persist and link all of the data, the discovery process is long and costly. The mission of AI-ready data should be to enable a virtuous circle of discovery fueled by the last great pattern detected. Organizations don't want anything to stand in the way of that kind of velocity.

Trust in data

AI and its capacity to generate advice free of self-interest could go a long way in restoring trust in industries that sorely need it – such as pharmaceuticals or the financial industry. But first, organizations will need to consider the quality of data, as proved through its lineage and also cognitive bias, which – albeit unintentional – is baked into many AI models and data.

The ultimate expression of AI is its capacity to learn – to ingest the knowledge of the expert and of the crowd alike, to extract knowledge from historical data as well as data generated in the moment. The deeper the data set, the better the training. Of course, that is only true if the data is accurate. To know that data is trustworthy, organizations need to verify the data's provenance and lineage. As more organizations tap into unstructured data, including data streaming from GPS systems, satellites and IoT devices, the

challenge to track the provenance and lineage of data grows significantly.

Blockchain technologies can trace the provenance and lineage of data in robust and reliable detail. This includes the source of data, the identities of those who've had access to it, and any changes made to data, along with when and why they occurred. It eradicates data drift.

Maintaining the lineage of data is one challenge; eliminating bias is another. Often, the bias may exist in the algorithmic model itself. The humans who create those models may expect, for example, that the best data to determine creditworthiness is past history. In reality, other factors may be equally or even more determinative.

To date, more than 180 human biases – any one of which can affect how humans make decisions – have been defined and classified. The sheer complexity of identifying and eliminating each piece of potentially biased data makes it an excellent candidate for automation. Organizations are learning to train the models themselves to recognize bias and automatically suppress it.

As the decisions supported and made by AI systems become more consequential, employees and consumers alike seek assurance that algorithmically derived answers can be trusted. Interpretable AI models can offer answers with evidence. The challenge is to make that evidence easy for humans to understand.

In the meantime, organizations, including IBM, stipulate the importance of transparency – the need to be very clear about the data being used to train machines. Bias is just one aspect to consider. Human error, manipulation or data drift, which causes data to become inaccurate or incomplete over time, are just a few of the factors that render data less than ideal for training.

Data rights

At present, incumbent organizations enjoy the data advantage. It is, however, contingent. Changes to regulations, consumer sentiment, and even new business models could shift data pools just as surely as they shift profit pools.

New regulations, such as the General Data Protection Regulation (GDPR) governing the European Union, are dictating massive changes to how companies handle data – and even their rights to keep it proprietary. In Europe, for example, banks are being forced to give away their proprietary advantage by sharing their customer data with organizations in other industries. Cut off from personal data by new regulations, revenues from targeted ads could decline. User or customer information once freely shared with suppliers and other business partners could be curtailed.

In many countries, where data privacy and personal data regulations are not as restrictive as in Europe, organizations have been at the forefront of developing business platforms based on more liberal access to and use of data. But even in countries with looser regulations or enforcement of them, the expectation is that restrictions will grow over time. For organizations that expect to operate globally, the strictest standard could become the default. In China, the government is leveraging its access to citizens' personal data to develop a social credit system, which it plans to implement by 2020. Such a system would measure the overall "trustworthiness" of its citizens from both an economic and social perspective. The plan has generated considerable debate.

Changing consumer sentiment about the ownership of data – including who has the right to profit from it – has become another unknown. Even outperforming organizations can't be complacent. Paradoxically, because they draw more heavily on customer data than their peers, they may be the most at risk from any new constraints on personal data.

Concerns about data privacy and security, as well as complex cross-border regulations, are already constraining the use of personal data. If users know that personally identifiable information is theirs to command, they may be more likely to share it. Benefits could accrue to those organizations that have moved beyond mere compliance. Organizations that have committed to transparency and good policy are likely to earn the customer's trust, and with it, a new data advantage.

Biographies

Alessandro Curioni

Alessandro Curioni received his undergraduate degree in Theoretical Chemistry and his PhD from Scuola Normale Superiore, Pisa, Italy. He started at IBM Research-Zurich as a PhD student in 1993 before officially joining as a research staff member in 1998. He is an IBM fellow, vice president of Europe and director of the IBM Research Lab in Zurich, Switzerland. In addition to leading the IBM Research activities in Europe, he is also responsible for the global research in IoT and Security. Curioni is a world recognized leader in the area of high-performance computing and computational science where his innovative thinking and seminal contributions have helped solve some of the most complex scientific and technological problems in healthcare, aerospace, consumer goods and electronics. He was a member of the winning team recognized with the prestigious Gordon Bell Prize in 2013 and 2015. His most recent position has been the head of the Cognitive Computing and Computational Sciences department. In 2017 he was named a member of the Swiss Academy of Engineering Sciences.

Gianni De Fabritiis

Gianni De Fabritiis graduated in applied mathematics (1997, University of Bologna) with laude. He went on to work at the Italian CINECA supercomputing center (1998-1999), then completed a PhD (2002) from Queen Mary University of London. After one year working in a start-up in business intelligence, he took a position at University College London (2003-2006) and later won a tenure track position at University Pompeu Fabra. In 2006, he founded Acellera UK with a mission to accelerate the transition towards computerized drug discovery. He is currently ICREA Research Professor in Barcelona and Associate Professor and head of the computational science laboratory at the University Pompeu Fabra. The group research interests are rooted in application of computation to solve real world problems where we view intelligence as a form of computation. The research lines of the group are based on physics-based simulations and machine learning/AI with the aim to reach human level intelligence and better rational design in specific narrow environments. One such environment is biomedicine where computational methods are applied to drug design, protein folding, etc.

Sean Ekins

Sean Ekins graduated from the University of Aberdeen; receiving his M.Sc., Ph.D. in Clinical Pharmacology and D.Sc. in Science. He then came to the USA and was a postdoctoral fellow at Lilly Research Laboratories, worked as a senior scientist at Pfizer and then Eli Lilly. He went on to join startup companies as associate director of Computational Drug Discovery at Concurrent Pharmaceuticals Inc. (now Allergan), vice president of Computational Biology at GeneGo (now Thomson Reuters) and CEO at Phoenix Nest. He is currently founder and CEO of Collaborations Pharmaceuticals, Inc. which is focused on using machine learning approaches for rare and neglected disease drug discovery. He is also entrepreneur in residence in the Eshelman School of Pharmacy at UNC-Chapel Hill. He is also on the SAB of the Pitt Hopkins Research Foundation. Since 2005 he has been awarded over 20 NIH/DOD grants as well as performs as a consultant on many others. He has authored or co-authored >300 peer reviewed papers, book chapters, and edited 5 books. For over 23 years he has been at the forefront of using software in drug discovery research. He has a passion for finding new collaborators and helping rare disease families find a pathway to a treatment for their children.

Ed J. Griffen

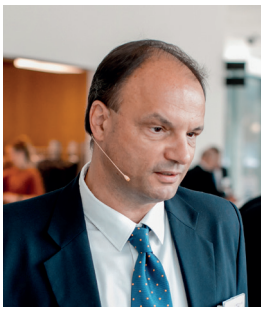
Ed J. Griffen obtained his Ph.D. from Imperial College, London, and undertook postdoctoral research at the University of Waterloo-Kitchener, Canada. He joined Zeneca Pharmaceuticals as a team leader in medicinal chemistry working in the CNS, infection, oncology, and chemical biology areas. During a secondment into the computational chemistry group, he co-developed matched molecular pair tools to quantify medicinal chemistry approaches. In 2012 he was a cofounder of MedChemica Ltd. He has taught medicinal chemistry courses at the University of Manchester and in France, and India. He is a named inventor on 16 patents and co-authored more than 22 articles, book chapters, and a textbook. He has worked on 4 projects that have entered clinical development. His interests are in developing data driven methods to enhance decision making and medicinal chemistry education.

Jürgen Schmidhuber

Since age 15, the main goal of Jürgen Schmidhuber has been to build a self-improving artificial intelligence (AI) smarter than himself, then retire. His lab's deep learning neural networks such as LSTM have revolutionized machine learning, are now on 3 billion smartphones, and used billions of times per day, for Facebook's automatic translation (2017), Google's speech recognition (since 2015), Apple's Siri & QuickType, Amazon's Alexa, etc. He also pioneered unsupervised adversarial networks, artificial curiosity and meta-learning machines that learn to learn. He is recipient of numerous awards, and chief scientist of the company NNAISENSE, which aims at building the first practical general-purpose AI. He is also advising various governments on AI strategies.

Alex Zhavoronkov

Alex Zhavoronkov holds two bachelor degrees from Queen's University, a master's in Biotechnology from Johns Hopkins University, and a PhD in Physics and Mathematics from Moscow State University. He is the Adjunct Professor of Artificial Intelligence at the Buck Institute for Research on Aging. He is the founder and CEO of Insilico Medicine, a leader in next-generation artificial intelligence technologies for drug discovery, biomarker development, and ageing research. At Insilico he pioneered the applications of generative adversarial networks and reinforcement learning for generating the novel molecular structures with the desired properties and generation of synthetic biological and patient data. He was the first to develop deep multi-modal predictors of age using multiple data types. He set up the R&D centers in 6 countries and launched multiple digital biomarker initiatives including Young.AI. Prior to founding Insilico Medicine, he worked in senior roles at ATI Technologies, NeuroG Neuroinformatics, YLabs.AI and established AgeNet.net competitions and Diversity.AI initiative. He is the co-founder and chief science officer of the Biogerontology Research Foundation, a registered UK charity focusing on age-related diseases. Since 2012 he published over 130 peer-reviewed research papers and 2 books. He serves on the editorial boards of some of the highest-impact journals in the field, and chairs the Annual Aging Research for Drug Discovery Forum and the Artificial Intelligence for Healthcare Forum at the Basel Life.



Andrea Danani

Andrea Danani graduated in Physics at ETH Zurich in 1987. After his PhD at SISSA/ISAS in Trieste in 1992, he was postdoctoral fellow at the Physics departments of Polytechnic University of Turin and University of Milan, and worked as scientific collaborator by the Swiss National Supercomputing Center (CSCS) in Manno. In 2001, Andrea Danani moved to SUPSI as scientific collaborator and teacher, where he received the Professor title in 2006. He is currently head of the Computational Biophysics Group at IDSIA, the Dalle Molle Institute for Artificial Intelligence affiliated to USI and SUPSI. Andrea Danani shows a long-standing expertise in molecular and multiscale modelling of several physical systems, such as many-electron systems and their phase transitions, polymer nanocomposites, biological systems in interaction with drugs and nucleic acids, macromolecules and nanoparticles for biomedical applications. In more recent years, he has applied classical and enhanced computational techniques to investigate protein conformational properties and protein-protein interactions responsible for their aggregation, with particular attention to related diseases such as Ataxia and Alzheimer. With similar approaches, he is also collaborating with pharmaceutical companies on Virtual Screening projects implementing standard docking techniques with machine learning.

Silvia Misiti

Silvia Misiti, MD Ph.D. In 2001 she started working for Sapienza University of Rome as a researcher in Endocrinology. In 2012 she moved to Lugano, where she lives and directs the IBSA Foundation for scientific research, a non profit organization founded by the pharmaceutical company IBSA, Institut Biochimique SA.

Her mission is to combine her great passion in scientific research with the promotion of different activities focused on innovation, education and dissemination, by collaborating with cultural and academic institutions. She is also Head of Corporate Communication & CSR for IBSA.

Martin O.

Vocal virtuoso, beatboxer, musical storyteller: Martin O. is all in one, in every way. Evenings go by in a flash as he invites his audience to dive into his playful virtuosity, his charm, humour, fantasy and, of course, his voice. He sings arias, yodels, raps, creates whole orchestras and bands, supplying his own sound effects. Martin O.'s witty dialogues in fancy languages are unmatched. His projections are fascinating, his interludes of song moving and his anecdotes amusing.

Martin O. grew up in a musical family. He started playing the violin and piano at an early age, and his talent for creating imagery entirely by voice was soon discovered. In 2007 he went on tour with his first program Dancing Voice, in 2011 he hit the stage with an all new show Cosmophon and in 2016 he presented his third program DER MaUSIKER (THE MoUSICIAN), where he walks in the footsteps of his great-great-uncle over the continents, across time and a variety of music styles. Martin O. has been awarded various national and international prizes in the "cabaret/chanson/song" category.

IBSA Foundation Papers

1 *Treatment of subclinical hypothyroidism in children, in women, and in adults*
May 10-12, 2013, Baveno-Stresa (also available in Italian)

2 *News for hypothyroidism*
September 27-29, 2013, Gubbio (also available in Italian)

3 *Stem cell therapy: hype or hope?*
March 29th, 2014, Lugano

4 *Metabolic diseases and tendinopathies: the missing link*
June 21st, 2014, Lugano

5 *Aging: is it a disease?*
September 27th, 2014, Frankfurt am Main

6 *Sexual health and sexual ecology*
May 9th, 2015, Milan

7 *Stress, inflammation and reproduction*
July 3rd, 2015, Siena

8 *Cancer immunology makes it to clinic: how cancer will be treated in the coming years*
September 26th, 2015, Lugano

9 *The thyroid… in the periphery!*
April 15th, 2016, Naples

10 *New technologies to treat neurodisorders: neuroprosthetics*
September 9th, 2016, Geneva

11 *Basic mechanisms of cancer immunotherapy*
June 16th, 2017, Lugano

12 *Female healthy aging*
September 7th, 2017, Zurich

13 *The new nutrition era: from molecular mechanisms to human health*
April 12th, 2018, Milan

14 *Sguardi scientifici sulle migrazioni/Scientific views on migrations*
October 13th, 2018, Lugano (bilingual Italian/English)

15 *Revolutionary therapies for cancer*
June 21st, 2019, Lugano

© copyright 2019 by Percorsi Editoriali of Carocci Publisher, Rome

Printed in September 2019 by Eurolit, Rome

Cover by Falcinelli&Co. / Stefano Vittori
Graphic design by Alberto Sala Design

Reproduction prohibited under the law
(Article 171 of the Law of 22 April 1941 no. 633)

Without proper authorization, you may not reproduce this volume even partially, by any means, including photocopying, even for internal or educational purpose.

Artificial Intelligence (AI) and Machine Learning (ML) are advancing rapidly and are being used increasingly across industries, including biomedical research and healthcare delivery. There are many potential benefits of applying AI techniques to improve the development of new molecules and the identification of new targets, cutting R&D costs and time.

This particular Forum is organized by IBSA Foundation for scientific research in collaboration with Istituto Dalle Molle di studi sull'Intelligenza Artificiale (IDSIA). The Forum brings together leaders in science and innovation to explore opportunities for AI and ML to accelerate medical advances in the biomedical research. It covers a range of issues including: the promise of integrating AI technology into healthcare, how it is being used in biomedical research, its potential for enhancing clinical care and scientific discovery.

Free copy • Not for sale